MEDICAL UNIVERSITY
OF VIENNA

EXPLORATORY ANALYSIS OF THE TRANSCRIPTOME OF MYELOPROLIFERATIVE NEOPLASMS FOR DIAGNOSTIC AND THERAPEUTIC APPLICATIONS

Doctoral thesis at the Medical University of Vienna

for obtaining the academic degree

Doctor of Philosophy

Submitted by

Fiorella Schischlik

Supervisor

Robert Kralovics, PhD

CeMM Research Center for Molecular Medicine

of the Austrian Academy of Sciences

Lazarettgasse 14, AKH BT 25.3

1090 Vienna Austria

Vienna, Submission Date: 19.01.2019

DECLARATION

Hereby, I list all the contributions for each section of this thesis. The author of this thesis performed the research in the laboratory of Robert Kralovics at CeMM, Research Center for Molecular Medicine of the Austrian Academy of Sciences in Vienna, Austria.

Results **section 3.1** and associated material and methods **section 5.1**, results **section 3.3** and associated material and methods **section 5.2** contain unpublished data and results. The author of this thesis analyzed the data, prepared the figures and has written these sections. J. Milosevic, E. Bogner, T. Klampfl and C. Cleary prepared libraries for next-generation sequencing. R. Jäger performed the targeted re-sequencing experiments.

Results **section 3.2** was submitted as a regular article to the journal *Blood* on the 28th of October 2018 and is currently under revision. The revised version of the manuscript was included in results **section 3.2**. The author of this thesis is the first author of the manuscript and has conceived and designed the experiments, analyzed and interpreted the data, prepared the figures and wrote the manuscript. R. Jäger and E. Fuchs performed the targeted re-sequencing experiments. E. Hug and F. Rosebrock performed the peptide exchange assays. M. Schuster and C. Bock have contributed with analysis of differential expression data. R. Holly performed the validation experiments for aberrantly spliced genes. J. Milosevic and E. Bogner have prepared libraries for next-generation sequencing and validated fusion candidates. B. Gisslinger, H. Gisslinger, M. Schalling, E. Rumi, D. Pietra, and M. Cazzola provided patient material and relevant clinical information. G. Fischer and I. Faé performed HLA typing experiments for *SF3B1* mutated patients. L. Vulliard and J. Menche have helped with analysis of pairwise associations of genes and diseases. T. Haferlach, M. Meggendorfer and A. Stengel provided patient material and performed FISH experiments. R. Kralovics had overall responsibility for the research, designed experiments and contributed to the writing of the manuscript.

Introduction **section 1**, aims **section 2** and discussion **section 4** were written by the author of this thesis. Images were conceptualized and drawn by the author. **Figure 2** was adapted from a review article written by the author of the thesis. The original figure was conceptualized and drawn by the author (*Schischlik, F., Kralovics, R., 2017. Mutations in myeloproliferative neoplasms – their significance and clinical use. Expert Rev. Hematol. 10, 961–973*).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

ABSTRACT

Myeloproliferative Neoplasms (MPNs) are stem cell derived hematological cancers characterized by a clonal hematopoiesis and an increase of differentiated blood cells of the myeloid lineage. The following thesis focuses on *BCR-ABL1*-negative MPNs, namely essential thrombocythemia (ET), polycythemia vera (PV), and primary myelofibrosis (PMF). Several phases of disease evolution in MPN are recognized. Patients often present with a chronic phase defined by a stable disease course and appearance of somatic mutations. Disease driving mutations in *JAK2*, *CALR*, and *MPL* are found in the majority of ET, PV, and PMF diseases and shape the phenotypic landscape of MPN patients. Patients without mutations in MPN driver genes are referred to as 'triple-negative' and represent a group of patients with unmet need for concise diagnosis. Patients can progress from a chronic to an accelerated phase or to a blast phase (secondary acute myeloid leukemia sAML) defined by an increase in genetic complexity and worse survival. Most therapeutic interventions treat patient symptoms and do not eradicate the MPN disease cell.

Whole transcriptome sequencing was performed on granulocytes RNA of 104 chronic MPN and 9 patients transformed to secondary AML. Several established and novel computational workflows were applied for determining clonality in granulocytes, calling and filtering fusion genes, single nucleotide variants (SNVs) and insertion and deletions (Indels), exploring aberrant splicing, and for discovery and systematic mining of neoantigens on our transcriptome dataset of MPN patients. The extent of granulocyte clonality was estimated in female MPN patients and thereby an alternative clonality assay was introduced, inclusive for all females and for the diagnostic benefit of triple-negative patients. The mutational landscape of the transcriptome of MPN patients was explored and characterized. A scarce number of non-recurrent fusions was found in chronic and secondary AML patients. MPN patients have a complex landscape of SNVs and indels with high frequency of *SF3B1* mutations in PMF patients. Analysis of splicing defects identify aberrant 3' splicing as the most common splicing alteration in *SF3B1* mutated PMF patients. A list of genes with putative protein altering defects caused by mis-spliced exons was described. Finally, all protein altering defects were collected and binding affinity was predicted for MHC class I molecules to 8, 9, and 10 mer peptides derived from altered proteins of different mutation classes. Putative neoantigens derived from *CALR* and *MPL* mutations were found. In general, 62% of MPN patients showed evidence of recurrent neoantigens providing a potential usage for targeted immunotherapy.

ZUSAMMENFASSUNG

Myeloproliferative Neoplasien (MPN) sind eine seltene Form von Blutkrebsarten, die ihren Ursprung in defekten Stammzellen haben. Einige Besonderheiten sind ihre klonale Blutbildung (Hämatopoese) und eine Überproduktion von ausgereiften Blutzellen der myeloiden Entwicklungslinie. Die vorliegende Doktorarbeit befasst sich mit den B*CR-ABL1* negativen MPN Krankheiten. Diese sind Essentielle Thrombozythämie (ET), Polycythämia vera (PV) und Primäre Myelofibrose (PMF). MPN Erkrankungen durchlaufen unterschiedliche Phasen. Meistens werden Patienten in einer chronischen Phase diagnostiziert, die sich durch einen stabilen Krankheitsverlauf definiert. In dieser Phase werden somatische Mutationen in den Genen *JAK2, CALR* und *MPL* in mehr als 90% von MPN Patienten gefunden. Diese sogenannten MPN spezifischen Mutationen tragen zu phänotypischen Ausprägungen der Krankheit bei. Patienten, die keine Mutation in diesen Genen aufweisen, werden als dreifach-negativ ('triple-negative') bezeichnet. Die diagnostische Aufarbeitung dieser Patientengruppe wird durch diese Eigenschaft erschwert. Einige Patienten können von einem chronischen zu einem progressiven Krankheitsbild ('accelerated phase') voranschreiten. Bei manchen Patienten kann zusätzlich eine sekundäre akute myeloide Leukämie AML ('blast phase') auftreten. In beiden Fällen führt dies zu einem Anstieg an genetischer Komplexität und einer deutlich reduzierten Überlebenswahrscheinlichkeit. Die Mehrheit der therapeutischen Maßnahmen behandelt die Symptome der Patienten und befasst sich nicht mit der Ursache der Erkrankung, nämlich der defekten Stammzelle.

In dieser Doktorarbeit wurde an einer Kohorte von 104 chronisch erkrankten MPN Patienten und 9 weiteren Patienten mit progressiver sekundärer AML, whole-transcriptome'-Sequenzierung an Granulozyten RNA von Patienten durchgeführt. Verschiedene rechnergestützte Verfahrensschritte ('workflows') wurden etabliert, um Klonalität in Granulozyten zu ermitteln, um Fusionsgene, Einzelvarianten, Insertionen und Deletionen in unseren Sequenzierdaten zu finden, um Defekte im Spleißen von Exonen zu detektieren und schließlich um potenzielle Neoantigene in unseren Daten zu identifizieren. Mit diesen workflows war es möglich einen neuen alternativen Weg zur Ermittlung von Klonalität in Krebszellen aller weiblichen MPN Patienten zu beschreiben. Insbesondere wird ein großer Nutzen für die Diagnose von dreifach-negativ Patienten erwartet. In chronisch sowie sekundär erkrankten AML Patienten wurde eine kleine Anzahl an nicht-rezidiven Fusionsgenen gefunden. Diese Doktorarbeit beschreibt eine komplexe Landschaft an Mutationen und eine hohe Häufigkeit an Mutationen im Gen

SF3B1 in Patienten mit PMF. Eine Analyse von Defekten im Exon-Spleißen von Patienten mit *SF3B1* Mutationen, definiert anomales 3'Spleißen als den am häufigsten auftretenden Defekt. Viele dieser Defekte führen zu potenziellen Veränderungen in der Aminosäuresequenz eines Proteins. Alle bisher genannten Mutationen wurden in dieser Arbeit gesammelt, um schlussendlich Neoantigene zu identifizieren. Die Bestimmung von Neoantigenen beruht auf rechnergestützten Vorhersagen von Peptiden und MHC-I Molekülen, die spezifisch für jeden Patienten sind. Damit konnten Neoantigene in Patienten gefunden werden, die von *CALR* und *MPL* mutierten Proteinen stammen. Zusammengefasst konnten Neoantigene in 62% aller MPN Patienten ermittelt werden, Dieses Ergebnis stellt eine vielversprechende Ressource für personalisierte Immuntherapie für MPN Patienten dar.

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Myeloproliferative Neoplasms (MPNs) encompass several heterogeneous diseases arising from somatic mutations in hematopoietic stem cells. According to the 2016 World Health Organization (WHO) classification of myeloid neoplasms, MPNs comprise seven groups of hematologic malignancies: Chronic myeloid leukemia (CML), chronic neutrophilic leukemia (CNL), polycythemia vera (PV), primary myelofibrosis (PMF) (prefibrotic and overt PMF), essential thrombocythemia (ET), chronic esosinophilic leukemia (CEL) and unclassifiable MPNs (MPNu) (Arber et al., 2016). In 1951, Dameshek was the first to group these diseases with overlapping clinical and laboratory features together. He introduced the term "myeloproliferative disorders" to describe CML, PV, ET, and PMF which are often referred to as the four "classic" MPNs (Dameshek, 1951). CML has a disease defining oncogenic fusion *BCR-ABL1*, caused by a translocation between chromosome 9 and 22. The following thesis will focus on novel methods and workflows for using transcriptomic sequencing data of *BCR-ABL1* negative MPN patients for diagnostic and therapeutic applications. Therefore, this introduction will briefly review in **section 1.1** the clinical features of PV, PMF, and ET patients as well as patients that progressed to acute myeloid leukemia (sAML) from a previous MPN. Furthermore, **section 1.2.1**, **1.2.3**, and **1.2.4** describe the genetic characteristics of these malignancies used for diagnosis and prognosis. Absence of genetic characteristics requires other criteria for diagnosis such as presence of clonal markers. The quest for biologic markers of clonality will be reviewed in **section 1.2.2** of the introduction. Finally, the last **section 1.3** will summarize current and future therapeutic strategies and their caveats as well as novel treatment opportunities in the field of cancer immunotherapy for MPN patients.

1.1 Clinical features and diagnostic criteria of myeloproliferative neoplasms

PV, PMF and ET (or collectively referred to as MPNs) are chronic hematologic malignancies characterized by a clonal hematopoiesis and excessive production as well as clonal expansion of terminally differentiated cells of one or more of the myeloid lineages (e.g. erythroid, granulocytic, and megakaryocytic) (Swerdlow et al., 2017). The MPNs are diseases of the elderly, with the highest incidence between 50-70 years of age and a worldwide annual incidence of 0.44-5.87 cases per 100.000 population (0.84, 1.03, 0.47 for PV, ET, and PMF, respectively) (Titmarsh et al., 2014) . Younger patients have been observed in cases with familial clustering of MPN suggesting underlying germline mutations (Bellanne-Chantelot et al., 2006). Evidence of germline predisposing factors influencing MPN pathogenesis have been reported, but these are rare and only explain a small proportion of familial MPN (Rumi and Cazzola, 2017). MPNs are often described as indolent cancers with an unobtrusive onset and relatively stable disease course. However, these malignancies have a tendency to progress in disease severity due to bone marrow failure caused by fibrosis, malfunctioning hematopoiesis, or transformation to acute myeloid leukemia (AML) characterized by an acute blast phase (Abdulkarim et al., 2009).

1.1.1 Polycythemia vera

Polycythemia vera patients have increased production of red blood cells (RBC) and therefore high proliferation of cells of the erythroid lineage, but also of granulocytes and megakaryocytes (panmyelosis) (Swerdlow et al., 2017). Increase of cell mass, in particular RBC mass, leads to thickening of the blood (increased blood viscosity) and is the major cause of symptoms related to PV, which are hypertension or vascular problems. Vascular abnormalities include venous or arterial thrombotic events (e.g. deep vein thrombosis, myocardial ischemia, or stroke) (Marchioli et al., 2005; Swerdlow et al., 2017). Other frequent symptoms include headache, visual disturbances, dizziness, paraesthesia, and pruritus. Additionally, patients often present with palpable splenomegaly and hepatomegaly (Elliott and Tefferi, 2005; Swerdlow et al., 2017). Diagnostic criteria require the inclusion of major and/or minor criteria (**Table 1**) (Arber et al., 2016; Swerdlow et al., 2017). A major criterion is the presence of a *JAK2*-V617F or *JAK2*-exon12 mutation which is found in up to 95% of PV patients (**Figure 1**) (Baxter et al., 2005; James et al., 2005; Kralovics et al., 2005; Levine et al., 2005). Two phases of PV are generally accepted, a polycythemic (PV) phase and a post-polycythemic

myelofibrosis (post-PV MF) or spent phase. While the PV phase is characterized by clinical features described in **Table 1**, the post-PV MF phase is associated with inefficient hematopoiesis resulting in cytopenia and anemia, bone marrow reticulin and collagen fibrosis, extramedullary hematopoiesis and hypersplenism (Swerdlow et al., 2017). Further progression to accelerated phase or MDS are possible and are accompanied with an increase of peripheral blast cells. A percentage of blast cells of $\geq 20\%$ requires a change of diagnosis to blast-phase post-PV MF, or formally acute myeloid leukemia (AML) (Arber et al., 2016).

Table 1: Diagnostic criteria for MPN according to the 2017 World Health Organization (Arber et al., 2016; Swerdlow et al., 2017). Abbreviations: P-PMF (prefibrotic/early myelofibrosis), O-PMF (overt myelofibrosis).

Criteria	PV[a]	ET[b]	PMF[c]	P-PMF	O-PMF
Major	1. **Hemoglobin** [>16.5 g/dL in men; 16.0 g/dL in women] or, **Hematocrit** [>49% in men and 48% in women] or, **increased red blood cell mass** (RCM)* [>25% above mean normal predicted value]	1. **Platelet count** ≥ 450*10^9/L	1. **Megakaryocytic proliferation** and atypia, without **reticulin fibrosis** grad >1,	x	x
			accopanied by reticulin and/or collagen fibrosis grades 2-3,		x
			accompanied by increased age-adjusted, **bone marrow cellularity, granulocytic proliferation**, and (often) decreased erythropoiesis	x	
	2. Bone marrow biopsy showing age-adjusted **hypercellularity** with trilineage growth (panmyelosis) including prominent ery-throid, granulocytic, and megakaryocytic proliferation with pleomorphic, mature megakaryocytes (differences in size)	2. Bone marrow biopsy showing **proliferation mainly of the megakaryocyte lineage** with increased numbers of enlarged, mature megakaryocytes with hyperlobulated nuclei. No significant increase or left shift in neutrophil granulopoiesis or erythropoiesis and very rarely minor (grade 1) increase in reticulin fibers	2. WHO criteria for *BCR-ABL1* positve chronic myeloid leukemia, PV, ET, myelodysplastic syndromes, or other myeloid neoplasms are not met	x	x
	3. Presence of *JAK2* **V617F** or *JAK2* **exon 12** mutation	3. WHO criteria for *BCR-ABL1* positive chronic myeloid leukemia, PV, PMF or other myeloid neoplasms are not met	3. *JAK2*, *CALR*, or *MPL* mutation or Presence of another clonal marker[d] or	x	x
			Absence of minor reactive bone marrow reticulin fibrosis	x	
			Absence of reactive myelofibrosis		x
		4. *JAK2*, *CALR*, or *MPL* mutation			
Minor	Subnormal serum erythropoietin (EPO) level	Presence of a clonal marker or Absence of evidence of reactive thrombocytosis	Anemia not attributed to a comorbid condition Leukocytosis ≥ 11*10^9/L Palpable splenomegaly Lactate dehydrogenase level above the upper limit of the institutional reference range	x	x
			Leukoerythroblastosis		x

a The diagnosis of PV requires either all 3 major criteria or the first 2 major criteria plus the minor criterion. Major criterion 2 (bone marrow biopsy) may not be required in patients with sustained absolute erythrocytosis, if major criterion 3 and the minor criterion are present. However, initial myelofibrosis (present in as many as 20% of patients) can only be detected by bone marrow biopsy, and this finding may predict a more rapid progression to overt myelofibrosis (post-PV MF)

b The diagnosis of ET requires that either all major citeria of the first 3 major criteria plus the minor criterion are met

c The diagnosis of prefibrotic/early myelofibrosis and overt myelofibrosis requires that all 3 major criteria and at least 1 minor criterion are met.

d In the absence of any of the 3 major clonal mutations, a search for other mutations associated with myeloid neoplasms (eg. ASXL1, EZH2, TET2, IDH1, IDH2, SRSF2, SF3B1 mutations) may be of help in determining the clonal nature of the disease.

1.1.2 Essential thrombocythemia

In essential thrombocythemia the main cell lineage involved is the megakaryocytic lineage. A persistent thrombocytosis (platelet count $\geq 450 \times 10^9$/L) in the peripheral blood is a major criterion for diagnosis. Other major criteria (summarized in **Table 1**) include bone marrow biopsies with occurrence of large megakaryocytes and presence of mutations in either *JAK2* (in 50-60% of ET patients), *CALR* (~30%), or *MPL* (~3%) (**Figure 1**) (Arber et al., 2016; Swerdlow et al., 2017). However, absence of strong disease

defining features, make exclusion criteria like presence of *BCR-ABL1* fusion, or evidence of reactive thrombocythemia necessary (Swerdlow et al., 2017). In particular triple-negative ET patients, which are negative for mutations in *JAK2*, *CALR*, or *MPL* (around 12% in ET), rely heavily on exclusion criteria for diagnosis (Cabagnols et al., 2015b; Harrison and Vannucchi, 2016; Milosevic Feenstra et al., 2016). ET is the most indolent of MPNs and more than 50% of patients are asymptomatic at the time of diagnosis. If symptoms are present, these usually manifest as vascular occlusions or bleeding mostly from mucosal surfaces. Mild splenomegaly (50%) or hepatomegaly (15-20%) can be present at diagnosis (Swerdlow et al., 2017). Similar to PV, ET patients can progress to myelofibrosis (post-ET MF) or acute myeloid leukemia in less than 5% of cases (Abdulkarim et al., 2009).
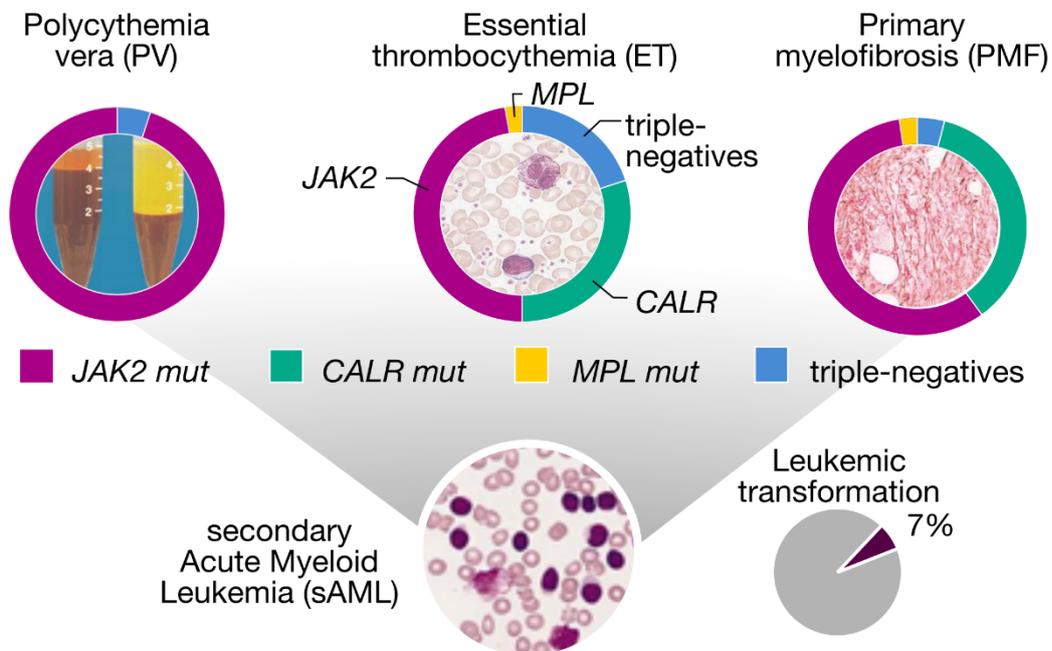
Figure 1: Ph-negative Myeloproliferative Neoplasms (MPN) encompass 3 disease subgroups with distinct phenotypic and clinical features. Images within donut chart were adapted from (Campbell and Green, 2006). Left image: PV is characterized by increase in hematocrit in the peripheral blood (test tube). Middle image: Peripheral blood smear of a patient with ET shows marked thrombocytosis. Right image: Bone marrow (reticulin stain) biopsy in a PMF patient with reticulin fibrosis. Lower image: Peripheral blood smear of a sAML patient with increase of immature blood cells (blasts). The overall percentage of transformation was estimated around 7% in a study of 795 MPN patients (Abdulkarim et al., 2009). However, risk for transformation is lower for PV and ET (1-3%) and higher for PMF (10-20%) and post-PV/ET MF (~50%) (Abdulkarim et al., 2009; Cerquozzi and Tefferi, 2015).

1.1.3 Primary myelofibrosis

Primary myelofibrosis (PMF) is characterized by expansion of cells mainly of the megakaryocyte but also the granulocyte lineage and fibrotic depositions in the bone marrow. Mutations in all three driver genes *JAK2* (40-60%), *CALR* (~24%), and *MPL* (~8%) are frequent. Triple-negative cases account for ~12% of PMF patients. (**Figure 1**). A prefibrotic/early PMF and an advanced or overt PMF phase is recognized (Swerdlow et al., 2017). However, this diagnostic dichotomy is better represented as a continuum where PMF patients follow a stepwise evolution and progression in clinical features and symptoms. In early stages, patients present with low grade or absent fibrosis (grade 0-1) and marked thrombocytosis. These are characteristics often associated with ET, therefore careful morphologic analysis of bone marrow histopathologic features (cellularity, megakaryocyte size, etc.) is of significant importance (given differences in prognosis and survival) to distinguish prefibrotic PMF from ET (**Table 1**) (Rumi et al., 2018; Swerdlow et al., 2017). The fibrotic stage in PMF is characterized by marked fibrosis (grade 2-3) in the bone marrow, osteosclerosis, leukoerythroblastosis in the peripheral blood, a decrease in platelet count and often patients present with splenomegaly (90%) and hepatomegaly (50%) (Swerdlow et al., 2017). Constitutional symptoms are frequent (~50%) and reflect the biological activity of the disease. Symptoms include among others fatigue, weight loss, and low-grade fever. In more advanced disease stages fibrosis can be accompanied by extramedullary hematopoiesis (myeloid metaplasia), primarily in the spleen followed by the liver and an increase of blast (CD34+) cells in the peripheral blood (Barosi et al., 2001). Same as in PV and ET, patients with more than 20% blast cells are classified as progression to acute leukemia (Arber et al., 2016; Swerdlow et al., 2017).

1.1.4 Other myeloid malignancies

MPNs are related to other myeloid disorders such as myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML).

MDS is a group of clonal bone marrow malignancies characterized by peripheral cytopenias (patients are anemic) due to ineffective hematopoiesis (Arber et al., 2016). Morphologic dysplasia of one or more cell types of the myeloid lineage are common (at least 10% are required for diagnostic purposes) and transformations to leukemia are observed in about 30% of affected patients (Abdulkarim et al., 2009; Swerdlow et al., 2017).

Classification of **AML** diseases is based on disease defining cytogenetic and molecular abnormalities. Chromosomal translocations t(9;11), t(8;21), t(15;17), inversions inv(16), and fusion genes *PML-RARA*, *RUNX1-RUNX1T1*, *CBFB-MYH1B* are frequent (Arber et al., 2016; Swerdlow et al., 2017). Clinically, the presence of more than 20% blast cells in the peripheral blood or bone marrow is required for diagnosis. AML can arise *de novo* or are caused by cytotoxic therapy (therapy-related AML). A secondary cause as a result of a previous chronic malignancy such as MDS or MPN (secondary AML - sAML) is possible. Survival and prognosis for secondary AML patients is dismal, with a median survival of less than 6 months (Kennedy et al., 2013) .

1.2 Mutational landscape and genetic basis of MPN diseases

Diagnosis of *BCR-ABL1* negative MPN diseases was radically improved with the identification of mutations in *JAK2*, *MPL* and *CALR* (MPN driver genes) that have been shown to promote the disease development of MPN (Vainchenker and Kralovics, 2017). Importantly, none of the mutations in MPN driver genes are specific for any MPN subtype and can therefore not be used to differentiate one MPN subtype from the other. Other somatic MPN-associated mutations have been described but they are not restricted to MPN diseases and lack mutual exclusivity to MPN driver mutations. However, many have important roles in prognosis, or have been associated with progression. Others have been linked to disease initiation providing a clonal advantage, or have shown to lead to phenotypic changes (Schischlik and Kralovics, 2017).

1.2.1 Somatic mutations in MPN disease driving genes *JAK2*, *CALR*, *MPL*

The ***JAK2*-V617F** was the first mutation to be associated with MPN pathogenesis and was discovered in 2005 by several groups (Baxter et al., 2005; James et al., 2005; Kralovics et al., 2005; Levine et al., 2005). The amino acid change V617F is caused by a point mutation in exon 14 of the *JAK2* gene and is able to activate three major cytokine receptors of the myeloid lineage, namely the erythropoietin receptor (EPOR), thrombopoietin receptor (MPL) and granulocyte colony-stimulating factor receptor (GCSFR) (Lu et al., 2005). The JAK2 protein is non-receptor tyrosine kinase and member of the JAK family tyrosine kinases which includes JAK1, JAK3 and TYK2. Cytokine binding to their respective receptors leads to a conformational change of the cytokine receptors and subsequent *trans*-phosphorylation of the JAKs. STAT proteins are recruited and are phosphorylated by the JAKs. STAT phosphorylation leads to their dimerization

and translocation to the nucleus to activate gene transcription for cellular pathways such as proliferation, differentiation and resistance to apoptosis (survival) (**Figure 2**).
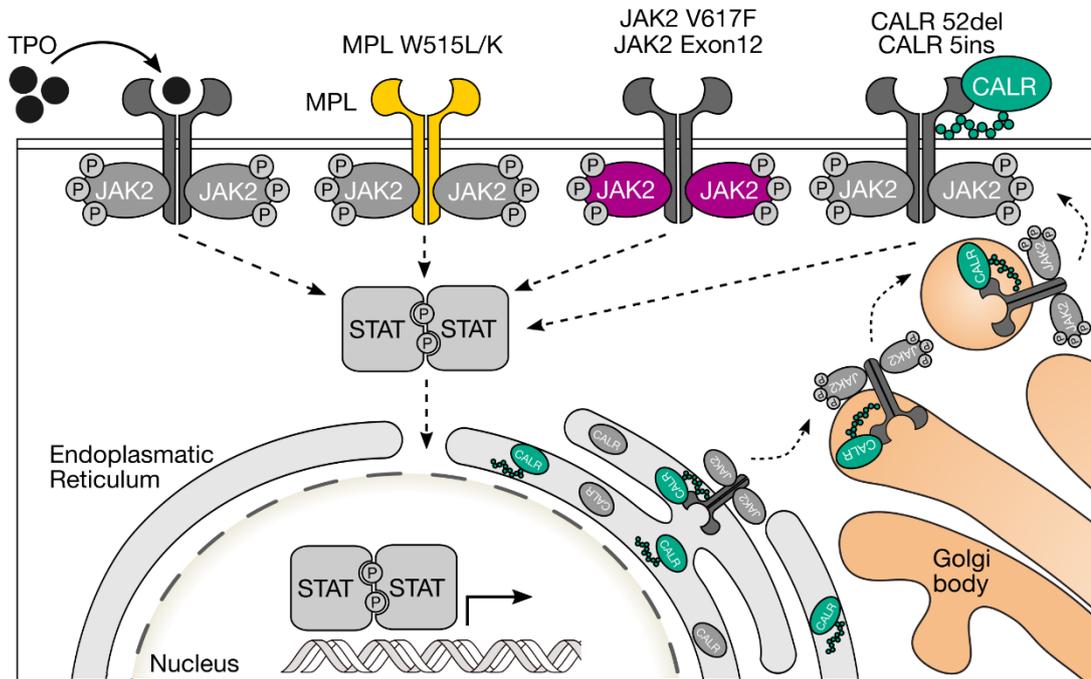
Figure 2: JAK-STAT signaling activation is a common feature in MPN pathogenesis. In canonical pathway, cytokine binding (e.g. thrombopoietin (TPO), erythropoietin (EPO)), induces a conformational change of cytokine receptors leading to activation of JAK proteins through *trans*-phosphorylation. Consequently, the JAK proteins phosphorylate the receptors and attract STAT molecules which are phosphorylated by the JAK proteins and form homodimers. STAT homodimers translocate to the nucleus and act as transcription factors for regulation of target genes. Mutations in *MPL*, *JAK2* and *CALR* lead to cytokine independent JAK-STAT signaling activation. Adapted from (Schischlik and Kralovics, 2017)

JAK family proteins comprise four domains, a tyrosine kinase (JH1) and a pseudokinase (JH2) domain, a FERM and a SH2 domain. All JAK proteins act through the JH1 domain, but the JH2 domain was shown to act as negative regulator of JAK protein function (Saharinen and Silvennoinen, 2002). The V617F mutation, which is located in the JH2 domain of the JAK2 protein leads to impaired and negative regulation of the JH1 domain and constitutively activated JAK2 kinase. Several mouse models for *JAK2*-V617F exist and are capable of mimicking several aspects of MPN pathogenesis and phenotypic features (J. Li et al., 2011). The mutant (or variant) allele burden of *JAK2*-V617F in MPN granulocytes ranges from detection threshold (as low as 1%) up to fully clonal (100%) (Rumi et al., 2014a). ET patients usually present with the lowest average burden and PV patients often have measured *JAK2*-V617F burden higher than 50%. Frequently, in PV patients and rarely in ET patients, higher burdens are accompanied by uniparental

disomies (UPDs) on chromosome 9p (Kralovics et al., 2002). 9pUPDs are among the most common chromosomal aberrations in MPN patients and target the *JAK2* gene, leading to a loss of wildtype allele and homozygous mutation status of *JAK2*-V617F. Interestingly, in a *Jak2*-V617F transgenic mouse model (designed to express *Jak2*-V617F in either low level or as high as endogenous wildtype *Jak2*) mice with low *Jak2*-V617F expression showed ET-like phenotypes, whereas high *Jak2*-V617F expression resembled clinically PV-like phenotypes (Tiedt et al., 2008). This and other studies modelling changes in burden (Akada et al., 2010), have demonstrated that differences in mutant gene dosage can explain some manifestation of MPN subtypes. However, it does not explain how one mutation can lead to three distinct MPN phenotypes.

Besides the *JAK2*-V617F in exon 14, *in frame* small insertions and deletions in exon 12 were found in low frequency in PV patients and have similar effects (Scott, 2011; Scott et al., 2007).

Mutations in the thrombopoietin receptor (*TPOR*) or myeloproliferative leukemia virus (***MPL***) are restricted to two hotspot regions on amino acid position W515 and S505 in exon 10 and are found in PMF (~5-8%) and ET (~3%) patients (**Figure 1**). The W515 is the most frequent and located in the cytoplasmic domain of the *MPL* receptor protein (after the transmembrane domain) and plays an important role in constitutive activation of the receptor and consequently cytokine-independent growth (shown in several Ba/F3 mouse and UT7 human cell lines) (Pardanani et al., 2006; Pikman et al., 2006). Tryptophan on position 515 is often replaced with leucine (W515L) or lysine (W515K). Amino acid changes to arginine (W515R), alanine (W515A), and glycine (W515G) exist but are less frequent (Defour et al., 2016). Mutations in *MPL* gene are usually heterozygous, but UPDs on chromosome 1p targeting the *MPL* gene occur often during disease progression (Rumi et al., 2013). The *MPL*-S505N mutation is very rare and was initially described as a germline mutation in familial thrombocythemia (Ding et al., 2004). However, somatic *MPL*-S505N mutations have been reported and emphasize the similarity of hereditary manifestations of thrombocythemia and true ET (Beer et al., 2008). In a study of 557 ET patients, mutations in *MPL* gene (12 patients in total) had a higher risk for fibrotic progression (Haider et al., 2016). In general, *MPL* mutated patients have a similar survival rate as *JAK2* mutated and worse survival than *CALR* mutated patients (Rumi et al., 2014b).

Calreticulin (**CALR**) mutations are exclusively insertions and deletions leading to shifts in the reading frame and found in ET (~30%) and PMF (~24%) patients (Klampfl et al., 2013; Nangalia et al., 2013). The most common (accounting for 85% of all mutations) is the type 1 (52 base pair (bp) deletion or *CALR*-del52) and the type 2 (5 bp insertion or *CALR*-ins5 – TTGTC nucleotide sequence). Other frameshift mutations (more than 35 different types have been described) are referred to as type 1-like or type 2-like, depending on their structural similarity to the type 1 and type 2 mutations (Eder-Azanza et al., 2014). *CALR* frameshift mutations are located on exon 9 and induce a +1 frameshift, which is the only reading frame known to be pathogenic. A frameshift in the last exon of the *CALR* gene leads to formation of a novel C-terminus and a change from negatively to positively charged amino acids. In addition, the protein sequence/motif KDEL is lost, which is an important signal for protein retention in the endoplasmatic reticulum (ER).

Early *in vitro* experiments have shown that overexpression of the *CALR*-del52 in Ba/F3 mouse cell lines leads to cytokine independent growth and activation of STAT5 (Klampfl et al., 2013). More recent publications could demonstrate that *CALR*-del52 and *CALR*-ins5 can activate MPL and consequently drive JAK-STAT pathway activation. Furthermore, it was shown that *CALR* mutants require a novel C terminus as previously described and MPL to display its oncogenic potential (Balligand et al., 2016; Chachoua et al., 2016; Elf et al., 2016; Marty et al., 2016; Nivarthi et al., 2016). It was shown that CALR mutant proteins as well as wildtype CALR proteins bind to MPL receptor in the ER. This interaction is intensified in the mutants with the novel C terminus (Araki et al., 2016; Chachoua et al., 2016; Elf et al., 2016). In the ER the mutant CALR remains bound to MPL and is then transported to the cell surface. However, the exact pathway from ER to the cell surface, as well as the exact cellular compartment of MPL activation is yet unclear. It is hypothesized that CALR acts as a chaperone gone rogue, constitutively activating MPL (**Figure 2**) (Staerk et al., 2006).

Differences in frequency of mutations have been reported between ET and PMF patients as well as distinct clinical feature between type 1 and type 2 mutations. For instance, in ET patients we find a relatively small difference (51% vs 39%) between type 1 and type 2 mutations, whereas in PMF the difference becomes considerably larger (70% vs 13%) (Cabagnols et al., 2015a). Differences in clinical features were also found between type 1 and type 2 mutations. Patients with type 1 mutations were more likely to have PMF and had significant reduced survival compared to type 1 mutations (Aylew Tefferi et al.,

2014). In addition, type 2 mutations are associated with higher platelet counts in ET patients (Ayalew Tefferi et al., 2014c).

1.2.2 Triple-negative MPN patients – Classification and the search for novel diagnostic markers of clonality

The term 'triple-negative' is originally derived from breast cancer patients for describing tumors negative for estrogen, progesterone receptor and *HER2* mutations. Since the discovery of the *CALR* mutations, triple-negative MPNs are referred to as MPN patients (mainly ET and PMF patients) with no mutations found in any of the three MPN driver genes *JAK2*, *CALR*, and *MPL*.

They have been described as a heterogeneous group of diseases with distinct prognosis in respect to survival. For instance, in ET patients, triple-negative patients have better survival compared to *JAK2* and *MPL* mutated patients (Rotunno et al., 2014). On the other hand, triple-negative PMF patients have the worst survival compared to patients with mutations in driver genes (Rumi et al., 2014b; Ayalew Tefferi et al., 2014a).

More recently, two studies identified non-canonical mutations in *JAK2* and *MPL* in 5% and 10% triple-negative ET and PMF patients, respectively (Cabagnols et al., 2015b; Milosevic Feenstra et al., 2016). In the Milosevic et al. study, 4 somatic (T119I, S204F, E230G, Y591D) and 1 germline mutation (R321W) in *MPL* gene were found, as well as non-canonical somatic and germline mutations in *JAK2*. In addition, many cases of ET and PMF patients with polyclonal hematopoiesis were detected using X-linked assays suggesting a different diagnosis than MPN. This study exemplifies the intricacies of proper diagnosis of triple-negative cases. The diagnostic workup of these disease entities requires in the "absence of any three major clonal disease drivers, the presence of other mutations (e.g. *ASXL1, EZH2, TET2, IDH1, IDH2, SRSF2, SF3B1*), to help define the clonal nature of the disease" (**Table 1**) (Swerdlow et al., 2017). If no clonal markers are to be found, absence of hereditary malignancies or reactive conditions (reactive thrombocytosis or myelofibrosis in ET and PMF, respectively) become vital exclusion criteria for ET or PMF diagnosis. However, it becomes evident, that the WHO criteria for diagnosis of triple-negative MPN are far from perfect to distinguish true ET and PMF from other polyclonal diseases.

Traditionally, determination of the clonal origin of a hematologic malignancy was accomplished with X chromosome inactivation (XCI) clonality assays. (Adamson et al., 1976; Fialkow et al., 1981). These assays take advantage of the unique mechanism of

XCI present in females. Females need to compensate for gene dosage on the X chromosome by epigenetically silencing either the maternal or paternal derived X chromosome (Beutler et al., 1962; Lyon, 1961) (**Figure 3**).
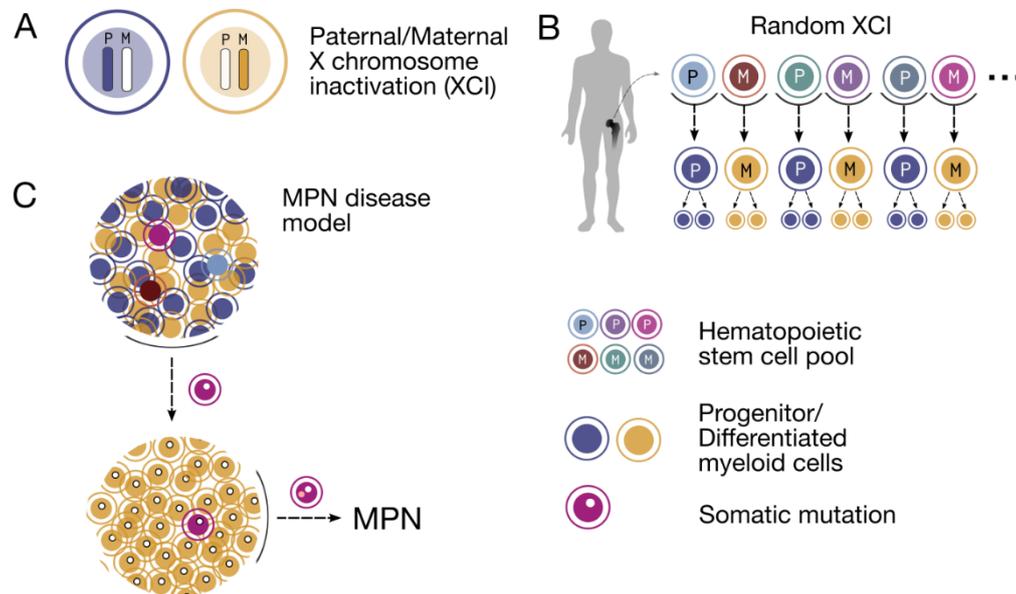
Figure 3: X-chromosome inactivation. (A) At an early stage in embryogenesis before hematopoietic lineage differentiation, either the maternal (M) or the paternal (P) derived X chromosome in each cell is inactivated. **(B)** The choice of chromosome is supposedly random and is maintained constant through generations of mitoses. Subsequently, healthy human females possess a hematopoietic stem cell pool, which is a mosaic of maternal or paternal silenced X chromosomes. The state of inactivation is propagated to differentiated cells of the lymphoid or myeloid cell lineage. **(C)** Somatic mutations in a single hematopoietic stem cell (HSC) can lead to clonal dominance (through selective advantage of MPN stem cells over normal counterparts), which is reflected through changes in proportion of maternal or paternal activated X chromosome. Secondary somatic mutations can drive the disease phenotype as exemplified in the disease model of MPN.

Three non-coding lncRNAs (*XIST*, its antisense partner *TSIX*, and *XITE*), located close to the center of the X chromosome, are known as major effectors of the XCI process. *XIST* is exclusively expressed from the inactive X chromosome and is responsible for initiation of silencing by coating the entire X chromosome (Penny et al., 1996). After XCI, genes located on the inactivated X chromosome are silent (not expressed). However, in humans 15-25% of genes have been reported to escape XCI ('escape genes'), either across tissues or in a tissue-specific manner (Berletch et al., 2011; Carrel and Willard, 2005; Slavney et al., 2015).

The process of XCI is supposedly random leading to a mixture of two cell populations where either the maternal or paternal X chromosome is marked for XCI (Augui et al., 2011). The XCI ratio reflects the proportion of cells where either the maternal or paternal

alleles on the X chromosome are active (**Figure 4A, 4B**). The distribution of XCI ratios follows a normal distribution in healthy females (**Figure 4A**). The XCI ratio can range from random XCI (50:50) to completely skewed XCI (100:0) (**Figure 4C**) (Amos-Landgraf et al., 2006; Szelinger et al., 2014).
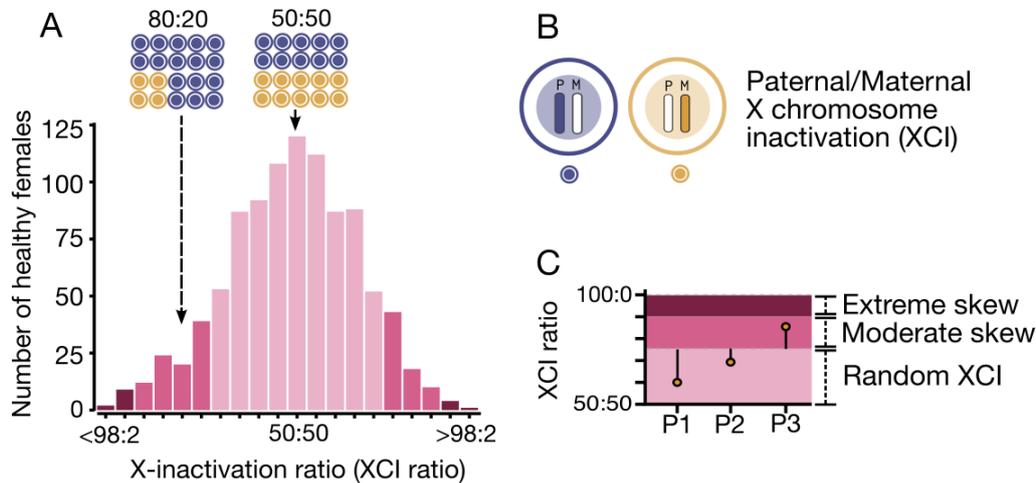
Figure 4: X-inactivation in healthy females. (A) The distribution of XCI ratios of 1005 healthy females is depicted. The XCI ratios were binned with increments of 5%. The distribution has a mean XCI ratio of 49:51 and a median of 50:50. The plot was adapted from Figure 1 in (Amos-Landgraf et al., 2006). **(B)** The XCI ratio represents the proportion of cells expressing genes from the paternal or maternal derived X chromosome. **(C)** Color shadings classify individuals with a random XCI (XCI ratio from 50:50 to 75:25), a moderate skew (XCI ratio from 75:25 to 90:10), or extreme skew (XCI ratio above 90:10).

Several assays have been developed to measure the extent of clonal cell populations. These function on the level of protein polymorphisms (G6PD isoenzyme) (Beutler et al., 1967), differential DNA methylation (*HUMARA*, *PGK*, or *HPRT*) (Allen et al., 1992) or on transcriptional level (Curnutte et al., 1992; Prchal and Guan, 1993) based on single gene or combination of several genes such as G6PD, MPP1/p55, IDS, FHL-1 (Chen and Prchal, 2007). In order to distinguish between the maternal or paternal origin of the X chromosome, it is essential that these surrogate markers of clonality are heterozygous. Conclusively, XCI clonality assays are restricted to informative females, heterozygous at these loci (Chen and Prchal, 2007).

1.2.3 Beyond MPN driver mutations

MPN patients often present with other somatic mutations besides *JAK2*, *CALR*, and *MPL*. Importantly, these mutations are not specific for MPN. In fact, they often occur in higher frequency in other myeloid malignancies like MDS and AML. In addition, these

mutations are not mutually exclusive to the above mentioned driver mutations, but for some genes differences in co-occurrences with driver mutations or disease subtypes have been reported (Lundberg et al., 2014a; Vainchenker and Kralovics, 2017). MPNs - similar to other hematological malignancies - have a low mutation frequency. The largest study used whole-exome sequencing on 143 MPN patients to provide an estimate on somatic mutation frequency of single nucleotide variants (SNVs) and small indels (Nangalia et al., 2013). PMF patients were shown to have the highest median number of 13 somatic mutations per patients compared to 6.5 for ET and PV patients. This observation has been correlated with PMF being the more progressed form of MPN diseases as well as reduced survival (Lundberg et al., 2014a; Vannucchi et al., 2013) and age (Grinfeld et al., 2018).

The Nangalia et al. study as well as other important publications in the field of MPN genetics, have helped to shape and characterize the genomic landscape of MPN patients using whole-exome and/or a combination of whole-exome and targeted sequencing on a restricted number of genes (Grinfeld et al., 2018; Klampfl et al., 2013; Lundberg et al., 2014b; Nangalia et al., 2013). Inclusion of germline controls enabled a true identification of somatic mutations. In MPN somatic mutations target genes involved in cytokine signaling (*LNK/SH2B3*, *SOCS*, and *CBL*), epigenetic changes (*TET2*, *DNMT3A*, *ASXL1*, *IDH1/2*, *PRC1 & 2*, *EZH2*), splicing (*SF3B1*, *SRSF2*, *U2AF1, ZRSR2*), and transcriptional regulators (*IKZF2, CUX1, FOXP1, ETV6, RUNX1, NF-E2*, and *TP53*). Several mutations within these genes are of prognostic value or have been assigned important roles in MPN pathogenesis (Schischlik and Kralovics, 2017). The following paragraphs will focus on a few important genes and pathways affected in MPN disease development.

Mutations in genes involved in **epigenetic changes** are among the most frequently mutated genes in MPN disorders (**Figure 5**). Among these *TET2* and *DNMT3A* function as key regulator of DNA methylation at CpG sequences (Tahiliani et al., 2009; Zhang et al., 2016). ***TET2 mutations*** are found in 13% in all subtypes of MPN (up to 20% in PV and PMF) patients, of which most are frameshift or stopgain mutations leading to putative loss-of-function of the protein (Delhommeau et al., 2009; Tefferi, 2010). Early studies showed that *TET2* mutations precede *JAK2*-V617F mutations and were important for establishing clonal dominance in hematopoietic stem cells (Delhommeau et al., 2009). However, *TET2* mutations have been found to either precede or follow *JAK2*-V617F mutations (Schaub et al., 2010). In fact, the order of acquiring mutations (*TET2*-first or *JAK2*-first), was shown to influence MPN phenotype. '*TET2*-first' patients were of older

age at diagnosis and had a reduced risk of thrombosis, resembling PV patients, whereas 'JAK2-first' patients were younger at diagnosis and had elevated risk of thrombosis, which are clinical features associated with ET patients (Ortmann et al., 2015).

Mutation frequency in **DMT3A** is ~10% in MPN patients (15% in PMF, 3-7% in ET and PV patients) and occur as gain of function (mainly on non-synonymous mutations in amino acid residue R882) or loss of function mutations (Abdel-Wahab et al., 2011; Stegelmann et al., 2011).

TET2 and *DNMT3A* mutations have shown to lead to expansion and increased self-renewal capacity in HSCs in mouse and human (Challen et al., 2012; Quivoron et al., 2011), emphasizing their role as disease initiators. This was further strengthened by the recent finding of somatic *TET2* and *DNMT3A* mutations (among other genes) in healthy individuals of advanced age. These studies found associations of clonal hematopoiesis in the elderly with higher risks of developing hematological cancers and decreased survival (Jaiswal et al., 2014; Lindberg et al., 2014; Xie et al., 2014). Interestingly, an earlier study had already described recurrent somatic *TET2* mutations in 5.5% of healthy individuals, of which some (n=7) were continuously followed for many years. One of these individuals developed ET and was tested positive for *JAK2*-V617F 5 years after follow-up (Busque et al., 2012).
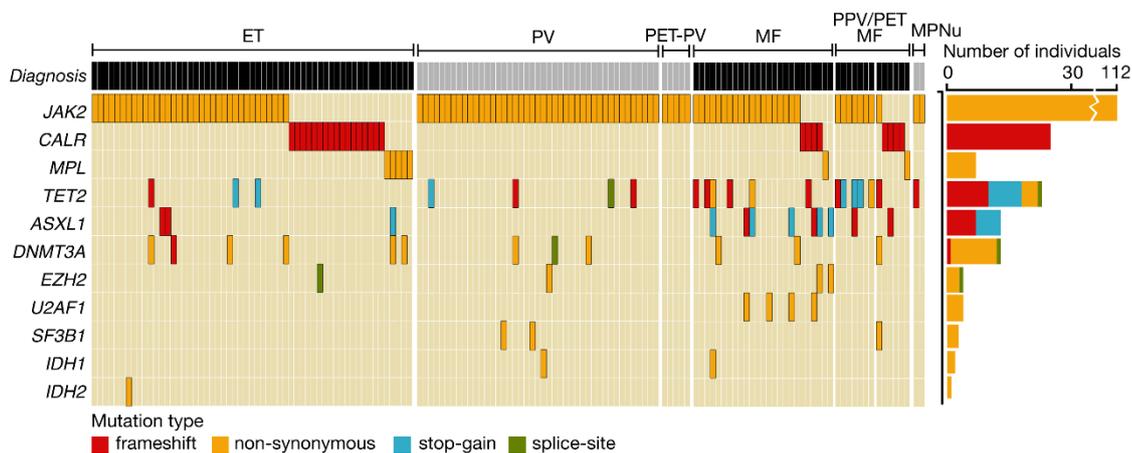
Figure 5: Analysis of mutation frequency as reported in (Nangalia et al., 2013). Data was extracted from Supplementary Methods. A total of 143 patients are depicted and selected genes are shown. Abbr.: PET-PV (Post-ET PV), PPV (Post-PV), MF (Myelofibrosis), MPNu (unclassifiable MPN).

Other genes frequently mutated in age-related clonal hematopoiesis were genes involved in the **splicing machinery** (**SF3B1**, **SRSF2**, and **U2AF1**). Recurrent mutations in genes involved in the spliceosome machinery were initially found in hematological cancers

(Papaemmanuil et al., 2011; Yoshida et al., 2011). Mutation frequencies varies among different disease classes, MDS with ring sideroblasts (85%) (Graubert et al., 2012; Papaemmanuil et al., 2011), chronic lymphocytic leukemia CLL (15%) (Quesada et al., 2012; Wang et al., 2011). In MPN, mutations in spliceosome genes are significantly enriched in PMF and post-MPN AML (3-7% in PMF for *SF3B1*, 9% in PMF and 19% post-MPN AML for *SRSF2*, 11% in PMF for *U2AF1*) (Lasho et al., 2012; Nangalia et al., 2013; Vannucchi et al., 2013; Zhang et al., 2012). Mutations in these genes are restricted to hotspot regions suggesting that these alterations lead to gain or altered functions of the protein. In *SF3B1*, mutations in the HEAT repeat domain of protein are frequently affected, among these are amino acid (aa) residue K700 and K666, in *SRSF2* the P95 aa residue and in *U2AF1* the S34 and Q157 aa residue are often mutated (Dvinge et al., 2016; Yoshida et al., 2011). Splicing factor mutations often occur as initiating mutations, early in disease development. For instance, in MDS analysis of clonal hierarchy, showed that *SF3B1* mutations are early events (Mian et al., 2015). In MPN, the clonal hierarchy of splicing factor mutations, has not clearly been resolved. A recent study in MPN patients suggested that *SF3B1* mutations are unlikely to occur as secondary events (Grinfeld et al., 2018).

Splicing factor mutations can be associated with good or worse survival. *SF3B1* mutations in CLL (Rossi et al., 2011) and *SRSF2* mutations in MPN (Zhang et al., 2012) are predictors of progression and worse outcome. However, in MDS and uveal melanoma (UV), *SF3B1* mutations are associated with good survival compared to *SF3B1* non-mutated patients (Harbour et al., 2013). Finally, in a study of 155 PMF patients, no worse survival in *SF3B1* mutated patients was found (Lasho et al., 2012).

SF3B1 is mutated in high frequency in hematologic but also in solid tumors. So far, in solid cancers, *SF3B1* mutations were identified in uveal melanoma (UV) (Furney et al., 2013), breast cancer (Maguire et al., 2015), malignant pleural mesothelioma (Bueno et al., 2016), bladder and pancreatic cancers (Biankin et al., 2012). Interestingly, the distribution of affected amino acid residues varies among different diseases. For instance, while the K666 amino acid is frequent in secondary and *de novo* AML, in other hematological malignancies the K700 is the predominantly affected amino acid residue (Dvinge et al., 2016). Whether this observation explains phenotypic and clinical characteristics of different diseases is less explored.

As previously mentioned, mutations accumulate in the HEAT repeat domain of the SF3B1 protein (Yoshida et al., 2011). The function of the HEAT domain has not fully been elucidated, therefore it is difficult to predict the mechanistic behavior of a mutant protein. It is known, that mutated SF3B1 affects 3' splice site recognition via the U2 complex. An *in silico* study by DeBoever et al. proposed a model in which mutant SF3B1 use cryptic 3' splice sites downstream of the branch point. The binding of the U2 complex to the branch point normally leads to 'steric occlusion' downstream of the branch point preventing recognition of cryptic 3' splice signals (AG). The authors in the study hypothesized that mutations in the HEAT domain cause an alteration in the sterically protected region and thereby allow for selection of alternative splice site (DeBoever et al., 2015). Further experimental evidence could show that the SF3B1 mutant U2 complex favors a different branch point than non-mutated SF3B1 (**Figure 6**). However, the mechanism of the preferential use of one branch point over another is unclear (Alsafadi et al., 2016; Darman et al., 2015).
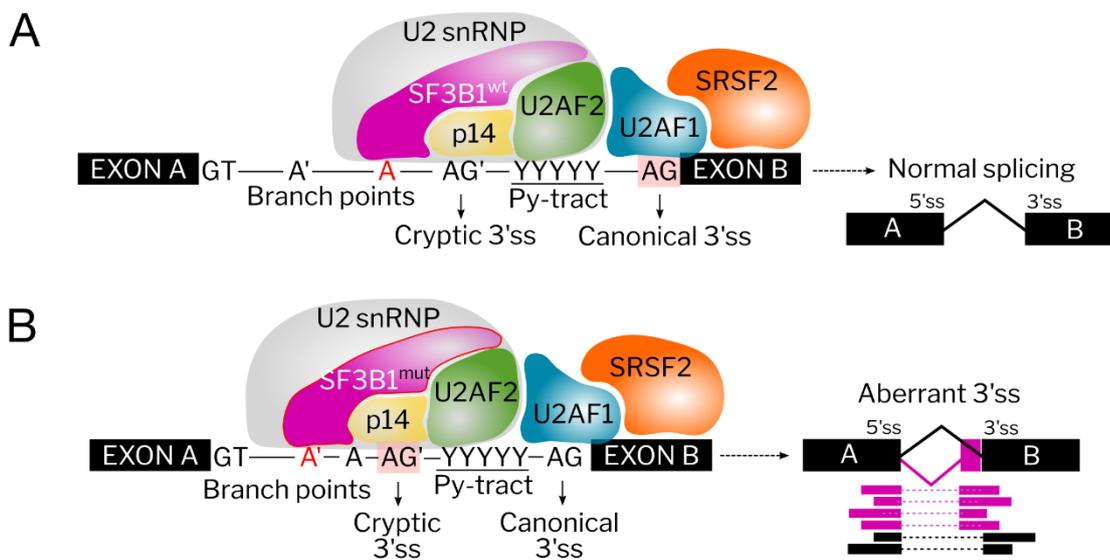
Figure 6: Model of aberrant 3' splicing of SF3B1 mutant protein. (A) Splicing is a tightly regulated program involving several large protein complexes and key sequence signals within the intronic sequence of the splice site. Towards the 3' splice site, the highly conserved AG nucleotides mark the beginning of the downstream exon (EXON B). Other splicing signals include a pyrimidine (Y) rich Py-tract and a branch point (A) recognized by the U2 snRNP complex. **(B)** Mutations in the HEAT domain of SF3B1 might induce a conformation change in the U2 snRNP structure, which now favors the alternate branch point (A') and recognition of cryptic 3' splice site. Alternate splicing can be quantified with RNA-seq, by calculating the ratio of reads mapping to the alternate versus the canonical splice site.

The above-mentioned studies were also among the first to explore the consequences of altered 3' splice site recognition in *SF3B1* mutated cell lines and patient samples. Differential junction analysis using RNA-seq data revealed up to several hundred mis-spliced junctions in MDS, CLL, UV and breast cancer. Furthermore, junction expression of *SF3B1* mutants varied among different cell types (cancer types) and specific amino acid change of the SF3B1 mutant protein. Although *SF3B1* mutations are frequently heterozygous and fully clonal, the ratio between the expression of the wildtype vs the mutant allele can vary extensively suggesting a preference for one or the other allele.

Whether individual or a combination of many genes affected by aberrant splicing are the downstream targets of *SF3B1* mutants is still a matter of investigation for many cancers. Darman et al. for instance proposed the mis-spliced *ABCB7* gene as a likely target in MDS with ring sideroblasts, given its role in iron metabolism (Darman et al., 2015). Another study investigated the role of mis-spliced *EZH2* in *SRSF2* mutated mice. They showed that *EZH2* mis-splicing triggers nonsense mediated-decay (NMD), leading to loss of *EZH2* which results in compromised hematopoietic differentiation and mimicking an MDS-like phenotype (Kim et al., 2015a).

Mutations in *SF3B1* are mutually exclusive to other mutations in genes of the spliceosome (*SRSF2*, *U2AF1*) and not tolerated in a homozygous state, suggesting a functional redundancy or a synthetic lethal relationship of splicing factor genes. A recent study has shed light on the mechanism of mutual exclusivity of two splicing factor genes *SF3B1* and *SRSF2*. The study showed a synthetic lethal interaction when both mutations were co-expressed. Additionally, they identified two mis-spliced genes (*MAP3K7* in *SF3B1* and *CASP8* in *SRSF2*) with convergent effects resulting in hyperactive NF-κB signaling (Lee et al., 2018).

1.2.4 Large aberrations, complex cytogenetic lesions and fusion genes in MPN

Classic cytogenetic karyotype analysis has still an important role in diagnosis of myeloid malignancies (Arber et al., 2016; Hussein et al., 2009). Karyotype analysis can identify chromosomal translocation and large aberrations such as inversions and deletions. However, the resolution is limited, and the exact genes involved in the rearrangements are not always known. Many of these rearrangements are disease defining and often the detection of such an event leads to direct classification of a hematological malignancy (e.g. *BCR-ABL1* in CML).

The emergence of SNP6.0 arrays allowed the identification of recurrent aberrations such as copy number changes (gains and deletions), and regions of loss of heterozygosity or uniparental disomies (UPDs) in MPN (Kralovics, 2008). Large aberrations are found in 62.5 % of MPN patients (at least one, but more in one patients occur) (Klampfl et al., 2011). Several studies showed, that the karyotype complexity, defined by the increasing number of aberrations detected with SNP6.0 arrays, does not differ between ET, PV, and PMF, but increases during disease progression to sMF or AML and are therefore of prognostic value (Klampfl et al., 2011; Milosevic et al., 2012). The most frequent lesions in chronic MPN were found on chromosome (chr) 9 (9p UPD & trisomy 9), chr 1 (1p UPD, 1q gain). Other frequent aberrations are del 13q, del 20q, del 4q, 14q UPD and 7q UPD (Klampfl et al., 2011).

SNP6.0 arrays allowed to map common deleted regions (CDRs) and identify the minimal deleted region in order to find the target gene of the aberration. Several transcription factors (*IKZF1, CUX1, CUX2, FOXP1, ETV6, RUNX1*) with putative involvement in MPN pathogenesis were mapped using this approach (Klampfl et al., 2011). However, CDR mapping with SNP6.0 arrays for identification of target genes has its limitations. First, it requires a critical number of patients with recurrent lesions. Second, only unbalanced chromosomal aberrations can be detected (lesions that induce copy number changes). Third, the low resolution renders an exact localization of the genomic breakpoints of the lesions prohibitive (Alkan et al., 2011).

Whole-exome and whole-genome sequencing have been used to detect small & large structural variations with base pair resolution (Alkan et al., 2011). While whole-exome sequencing can only be used for certain classes of lesions (unbalanced lesions and lesions where the breakpoint lies within the exonic region), whole-genome sequencing requires extensive sequencing, computational analysis and is costly for screening a large cohort of patients (Taylor and Ladanyi, 2011; Welch et al., 2011).

Gene fusions arising from chromosomal aberrations can act as tumor-specific biomarkers and are of tremendous therapeutic value for targeted precision medicine such as Imatinib, which is an inhibitor for the *BCR-ABL1* fusion gene in CML – a subtype of MPN. In myeloid malignancies (**Figure 7**) as well as in solid tumors, gene fusions are frequent disease drivers (Mertens et al., 2015; Mitelman et al., 2007; Yoshihara et al., 2014).
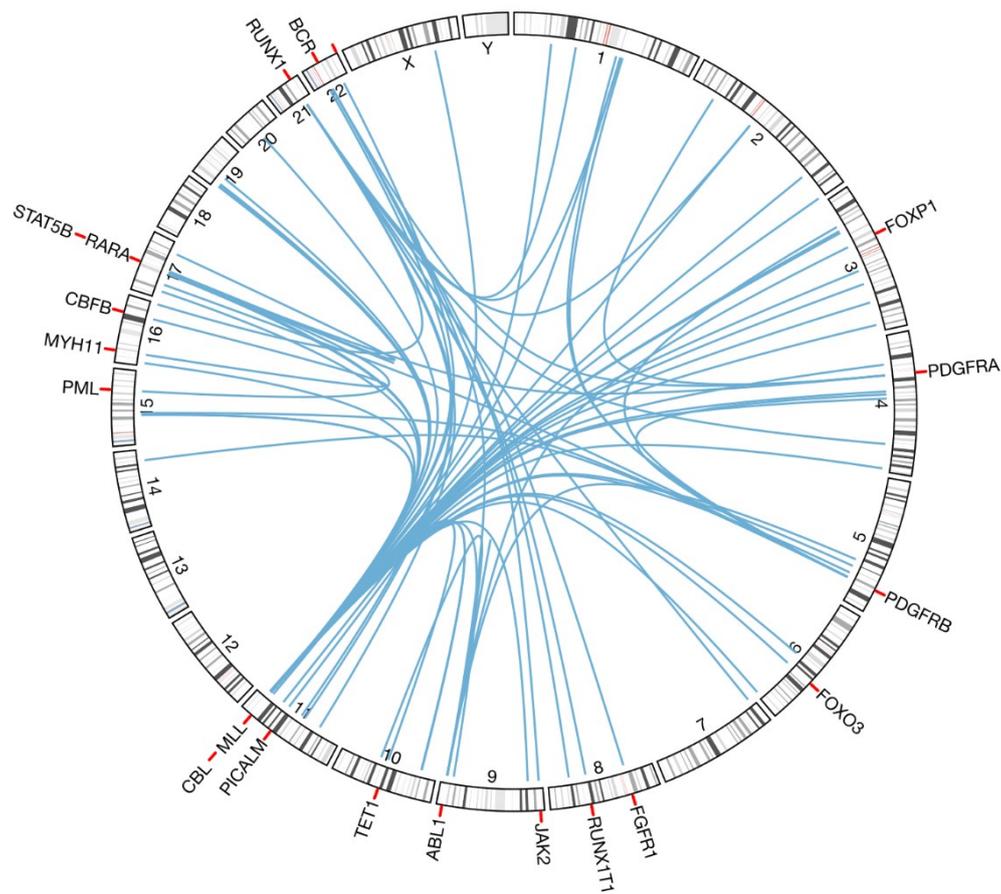
Figure 7: Complex landscape of fusion gene events in myeloid malignancies. Circos plot shows the analysis of 1280 fusion events in myeloid malignancies. Lines (=fusion events) connect two genes located between or within chromosomes. 1082 translocations, 68 insertions, 48 inversions, and 26 deletions leading to fusion events were analyzed. Only unique fusion events (n=241) are depicted. AML, CML, MDS, and chronic myelomonocytic leukemia (CMML) patients were included in this analysis. Selected gene names forming part of fusion events are displayed around the outer circle of the graph. Raw data was extracted from *The Mitelman Database of Chromosomal Aberrations and Gene Fusions in Cancer* (Mitelman et al., 2013).

RNA-sequencing has emerged as cost-effective and unbiased method for their systematic discovery (Cieślik and Chinnaiyan, 2017). Since the emergence of next-generation sequencing several software packages have been released with the aim to detect fusion genes with RNA-seq data (**Table 2**) (Wang et al., 2013). Although these tools have been successfully used in numerous publications for fusion gene discovery, if benchmarked against each other, the overlap between detected fusions is small (Abate et al., 2012; Carrara et al., 2013; Kumar et al., 2016). The reason is partly explained by the heterogeneity in tool specific implementation of alignment strategies (**Table 2**), the usage of reference genomes (whole genome, transcriptome) and an extensive variety of filters introduced to reduce the number of false positive fusions. Other factors of variability

include the method of sequencing (paired-end versus single end), the sequence coverage, the size of the insert sequence and the read length.

Table 2: List of software published for fusion gene detection. Tophat-fusion, deFuse and SOAPfuse were used for fusion detection in the results section of this thesis. Software packages were published for Tophat-fusion (Kim and Salzberg, 2011), deFuse (McPherson et al., 2011a), SOAPfuse (Jia et al., 2013), Fusioncatcher (Nicorici et al., 2014), FusionMap (Ge et al., 2011), TransABySS (Robertson et al., 2010), EricScript (Benelli et al., 2012), FusionQ (Liu et al., 2013), ChimeraScan (Iyer et al., 2011), ShortFuse (Kinsella et al., 2011), FusionHunter (Y. Li et al., 2011), MapSplice (Wang et al., 2010b), FusionFinder (Francis et al., 2012), Bellerophontes (Abate et al., 2012), SnowShoes-FTD (Asmann et al., 2011), FusionSeq (Sboner et al., 2010), Comrad (McPherson et al., 2011b), PRADA (Zheng et al., 2012), FusionAnalyser (Piazza et al., 2012), BreakFusion (Chen et al., 2012), EBARDenovo (Chu et al., 2013); N/A – not available

Fusion gene detection software	Software version	Read aligners	Library format
Tophat-fusion	2.0.8 release 26-2-2013	tophat (mod. bowtie)	RNA-seq
deFuse	release 2013-02-08	bowtie, BLAT	RNA-seq
SOAPfuse	1.25 release 04-12-2013	SOAP2	RNA-seq
Fusioncatcher	0.97 release 4-9-2013	bowtie, BLAT	RNA-seq
FusionMap	release 2013-02-01	modified GSNAP	WGS, RNA-seq
TransABySS	1.4.4 release 09-10-2012	GMAP, BWA	RNA-seq
EricScript	0.4.0 release 2013-3-20	BWA, BLAT	N/A
FusionQ	1.0 release 2013-06-15	tophat, cufflinks	RNA-seq
ChimeraScan	0.4.5 release 2012-2-25	bowtie	RNA-seq
ShortFuse	0.2 release 19-5-2011	bowtie	RNA-seq
FusionHunter	1.4 release 2012-06-09	bowtie	RNA-seq
MapSplice	1.15.1 release 2011-09-15	N/A	RNA-seq
FusionFinder	1.2.1 release 29-07-2012	N/A	RNA-seq
Bellerophontes	0.3.5 release 03-07-2012	N/A	N/A
SnowsShoes-FTD	release 2012	BWA	RNA-seq
FusionSeq	0.7.0 release 2011-05	bowtie, BLAT	RNA-seq
Comrad	0.1.3 release 2011-06-22	bowtie, BLAT	WGS, RNA-seq
PRADA	release 2012-11-14	N/A	RNA-seq
FusionAnalyser	release 2012-09	BWA	RNA-seq
BreakFusion	1.0 release 01-05-2012	BLAT	RNA-seq
EBARDenovo	1.2.2 release 2013-04-04	assembly	N/A

Like SNVs or Indels, fusion genes may have distinct functional outcomes. One possible scenario is an **enhanced overexpression of an oncogene**. For instance, the IgH-*MYC* fusions are formed through promoter swapping and activates *MYC* expression, which is under the control of a promotor of a highly expressed gene (e.g. Ig or T cell receptor).

Formation of **novel chimeric proteins** are also common. *BCR-ABL1* and *NPM1-ALK* are prominent examples of fusion genes, where an 'activator' gene like *BCR* or *NPM1* drives constitutive activation of a tyrosine kinase (*ABL1* and *NPM1*). Fusion events involving

the *JAK2* locus represent similar examples of an activated kinase. The *JAK2* gene can form fusions with several partner genes such as *RUNX1, ETV6, PCM1, BCR* and other genes (Bousquet and Brousset, 2006; Griesinger et al., 2005; Lacronique et al., 1997; Peeters et al., 1997) . *PCM1-JAK2* fusions were found in MDS/MPN, AML and T-cell lymphomas and are associated with myelofibrosis and eosinophilia and myeloproliferation (Hoeller et al., 2011).

Finally, fusion of two genes can lead to shifts in the reading frame (**Figure 8**). This is achieved either through introduction of novel nucleotides, fusion of non-matching reading frames of two exons, or fusion of exons placed in opposite transcriptional orientation.
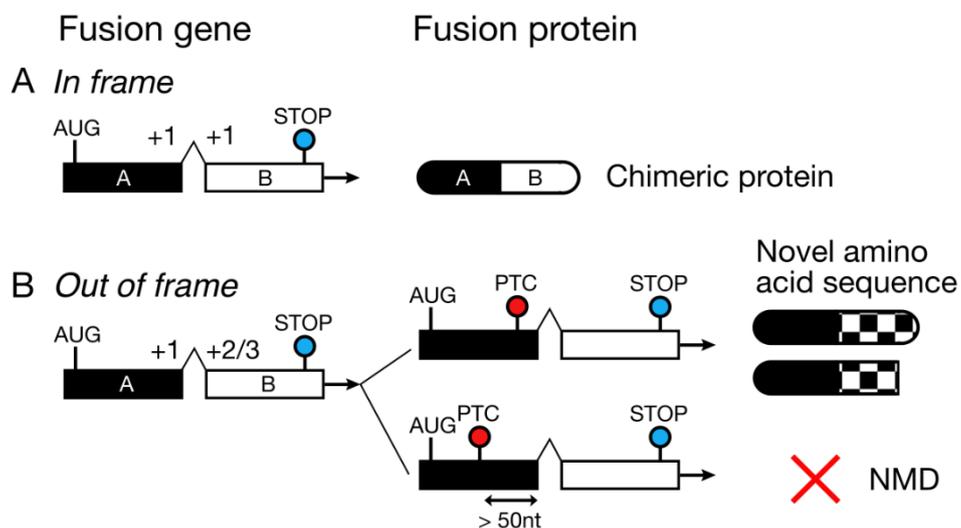
Figure 8: Schematic model of fusion protein formation. (A) Chimeric fusion proteins are formed through fusions of two genes in the correct frame of the exons located at the fusion breakpoint. **(B)** In frequent cases, two exons from different genes are fused *out of frame*. The result is a shift in the reading frame and often lead to the introduction of premature stop codons (PTCs). Consequently, fusion proteins have a novel C terminal end. Alternatively, the fusion transcript is degraded via nonsense mediated decay (NMD).

A possible scenario is the **inactivation of one or both genes** involved in the fusion. The landscape of inactivating fusions is certainly less explored. There is bias towards fusion forming chimeric proteins, since these form better drug targets. Nevertheless, inactivating fusion can be equally important to explain tumor formation. For instance, loss of tumor suppressor function was shown for fusion gene *PPP2RA-CHEK2* (Jin et al., 2006). Alternatively, *out of frame* fusions do not necessarily lead to degradation of the fused transcript. It is possible that out of frame transcripts are translated and create aberrant C-terminals of the chimeric protein. In a study by Rodriguez-Perales et al. a truncated RUNX1 fusion protein, expressed an altered C-terminal, which was further shown to

interfere with the wildtype RUNX1 protein in a dominant-negative and dose-dependent manner (Rodriguez-Perales et al., 2015).

In MPN, isolated cases of fusions in PV and PMF have been reported (Murati et al., 2006; Tanaka et al., 2010; Tiziana Storlazzi et al., 2014). Two patients tested negative for *JAK2-V617F*, *JAK2*-exon12 and *MPL* mutations, suggesting an accumulation of rare events in triple-negative patients. Importantly, at the time of publication, the *CALR* mutations had not been described, therefore a re-evaluation of the true nature of these triple-negative MPN cases is necessary. Large patient cohorts were not screened to determine recurrent fusion events in MPN. Co-occurrences of *BCR-ABL1* fusions with MPN driver mutations (*JAK2* and *CALR*) are rare in MPN patients but exits (Bornhäuser et al., 2007; Seghatoleslami et al., 2016; Yamada et al., 2014; Zhou et al., 2015). Patient often present with atypical phenotypes resembling CML. This emphasizes the complexity of proper diagnosis in patients with overlapping disease features (Kwong et al., 1996; Rice and Popat, 2005).

1.3 Therapeutic management of MPN patients

1.3.1 Current therapies for MPN and their limitations

Current drug therapy for MPN patients are not curative and are mainly used to prevent complications and treat symptoms arising from thrombosis, bleeding, anemia, and spleno-megaly (Vannucchi and Harrison, 2017). A risk-adapted treatment model exists for PV, ET, and PMF and is applied for each patient (Tefferi et al., 2018b, 2018c, 2018d).

For ET, a primary goal is to prevent thrombotic complications. Traditional risk stratification recognizes two parameters: advanced age (>60 years) and a previous history of thrombosis. Recently, the presence of a *JAK2* or *MPL* mutations has been identified as an independent risk-factor (Barbui et al., 2015). Patients who present with one of these risk factors are considered "high-risk" and in the absence "low-risk" patients. For low-risk patients, a conservative treatment is preferred, which can be either observational or low doses of aspirin. High risk patients are treated with cytoreductive therapy, usually hydroxyurea as a first-line therapy and pegylated interferon-α (INF-α), busulfan, anagrelide or pipobroman as second line therapy (Tefferi et al., 2018c). Of these, INF-α is the preferred drug and pipobroman the least recommended due to its association with leukemic progression in PV patients (Kiladjian et al., 2011; Tefferi et al., 2013).

Independent risk factors for overall survival in PV patients were identified as advanced age (>61 years), leukocyte count higher than 10.5×10^9/L, venous thrombosis and abnormal karyotype (Bonicelli et al., 2013; Tefferi et al., 2013). Low risk PV patients are treated with phlebotomy and low dose aspirin, whereas high risk patients receive similar treatment as high-risk ET patients (Tefferi et al., 2018d).

Several prognostic models for PMF patients have been evaluated and are in use (Tefferi et al., 2018b). The widely used dynamic international prognostic scoring system (DIPSS) plus model, has 8 risk factors: advanced age (>65 years), hemoglobin less than 10g/dL, leukocyte count more than 25×10^9/L, circulating blasts of more than 1%, the presence of constitutional symptoms, unfavorable karyotype, red cell transfusion dependency and a platelet count less than 100×10^9/L (Gangat et al., 2011). The presence of 0,1, 2, equal or more than 3 risk features defines in which of the 5 risk categories (low to high) a patient corresponds to. Several additional independent risk factors have been identified for PMF patients, these include among others, the absence of type 1 or type1 like *CALR* mutations, presence of mutations in the gene *ASXL1, EZH2, IDH1/2*, and in splicing factor *SRSF2* and *U2AF1*-Q157 (Tefferi et al., 2018a; Ayalew Tefferi et al., 2014b; Vannucchi et al., 2013). Low risk patients are often asymptomatic and generally closely monitored for changes in disease progression. Allogeneic stem cell transplantation (ASCT) is the only curative treatment for PMF patients. Due to high treatment related mortality for ASCT, it is only recommended for a minority of younger and high-risk patients (Ballen et al., 2010). The first medication approved by the Food and Drug Administration (FDA) to treat intermediate and high-risk MF patients was ruxolitinib, a non-mutant specific JAK2 inhibitor in 2011. Ruxolitinib was shown to significantly improve blood counts, splenomegaly and constitutional symptoms (Deisseroth et al., 2012). Other therapeutic drugs under investigation include immunomodulating agents such as thalidomide and lenalidomide, histone deacetylase (HDAC) inhibitors, telomerase inhibition (imetelstat), agents targeting fibrosis (e.g. anti TGF-β and anti-LOLX2) and pegylated INF-α (Shreenivas and Mascarenhas, 2018).

1.3.2 Therapeutic targeting of the MPN stem cell clone

MPN diseases are essentially stem cell disorders, originating from rare self-renewing stem cells. Therefore, eradicating the MPN disease stem cell is likely to lead to successful cure and patient treatment. However, even the most promising targeted therapies directed against leukemia stem cells in CML, fail to eliminate all stem cells in the majority of

patients, partly owing to their quiescence compared to progenitor cells (Gallipoli et al., 2011; Holyoake et al., 1999).Currently, besides ASCT, only few agent have shown to have an effect on the MPN disease stem cell clone quantified as reduction of mutant allele burden in MPN driver genes.

In a COMFORT-I study, a phase III trial for PMF patients, only 12% of patients treated with ruxolitinib had a more than 50% decrease in *JAK2*-V617F burden and less than 2% had a complete molecular remission. The duration of treatment had an influence on treatment success for ruxolitinib (Deininger et al., 2015). A second phase III study for MF patients retrospectively analyzed *CALR*-mutant patients treated with ruxolitinib. *CALR* mutant allele burden did not show a significant change after a median of 60 weeks treatment (Guglielmelli et al., 2016). In a small, prospective study, *JAK2*-V617F mutated PV or ET patients treated with hydroxyurea showed in 57% of patients a partial molecular response (Besses et al., 2011).

INF-α treatment is the only agent, which has consistently shown to reduce *JAK2*-V617F and *CALR* mutation burden in PV and ET patients (Gisslinger et al., 2015; Kiladjian et al., 2008; Quintás-Cardama et al., 2009; Them et al., 2015; Verger et al., 2015). In some patients a reversion from clonal to polyclonal hematopoiesis was observed (Liu et al., 2003; Massaro et al., 1997). However, resistance to INF-α therapy has been reported in patients where *JAK2*-V617F was not the initiating event or after cessation of treatment (Ishii et al., 2007; Kiladjian et al., 2010; Turlure et al., 2015). In addition, INF-α is not well tolerated in patients given its high toxicity rate (Soret et al., 2016).

1.3.3 Immunotherapy – emerging therapies for MPN patients

Immune-based therapies against cancers use the body's immune system to elicit an anti-tumor response and combat cancer. The cancer immunotherapy field has experienced tremendous advancements in the past years. Most notably, usage of monoclonal antibodies (mAbs), advances in immune checkpoint blockade therapy (CPB), chimeric antigen receptor (CAR) T cell as well as adoptive T cell therapy (ACT) and administration of cancer vaccines (Zhang and Chen, 2018).

A commonality between these methods is the requirement to identify optimal antigens or epitopes to target (Hu et al., 2017). Tumor antigens can be broadly categorized as tumor-associated antigens (TAAs) or tumor-specific antigens. Examples of TAAs include antigens, which are overexpressed or expressed higher in cancer tissue compared to

normal cells or antigens playing a role in tissue differentiation (Novellino et al., 2005). The lack of tumor specificity for TAAs increases the risk of inducing autoimmunity against the normal cells (Dudley et al., 2002). In addition, patients are likely to lack high-affinity T cells as a result of negative selection in the thymus (Theobald et al., 1997).

Tumor neoantigens on the other side arise through somatic mutations and were therefore considered along with oncogenic viral antigens as tumor specific and consequently ideal targets for ACT, CPB and therapeutic cancer vaccination (Hu et al., 2017; Schumacher and Schreiber, 2015; Shigehisa Kitano et al., 2015). However, their systematic discovery only became feasible with the advent of next-generation sequencing technologies. In addition, advances in the development of software to predict binding affinity of MHC molecules to epitopes enables the identification of putative neoantigens for each patient individually (Snyder and Chan, 2015; Zhang et al., 2011). Several studies have shown that neoantigens are crucial targets of anti-tumor T cell response. First, it was shown, that the number of neoantigens (or neoantigen load) correlates with better survival in patients (Brown et al., 2014). Second, CD8[+] and CD4[+] T cell populations specific for neoantigens, expand as a result of anti-tumor immune response (Linnemann et al., 2015; Rajasagi et al., 2014; Robbins et al., 2013; van Rooij et al., 2013). Finally, several *in vitro* and *in vivo* studies demonstrated the capability of neoantigen specific cytotoxic T-lymphocytes (CTLs) to effectively kill tumor cells (DuPage et al., 2012; Gubin et al., 2014; Kreiter et al., 2015; Tran et al., 2014).

Computational workflows to allow for quick identifications of personalized candidate neoantigens have been established. A typical workflow involves whole-exome sequencing of tumor and normal tissue to identify somatic mutations, followed by transcriptome sequencing of the tumor tissue to confirm expression of the mutated genes. Finally, mutations are translated to amino acid sequences and virtual peptide libraries of 8-11 mers in size are generated. The peptides are selected based on their probability to bind with high affinity to each individual's MHC molecule using prediction software such as NetMHCpan (Gross et al., 2014; Matsushita et al., 2012; Rajasagi et al., 2014; Robbins et al., 2013; van Rooij et al., 2013). Three studies could evaluate the immunogenicity, safety and feasibility of neoantigen cancer vaccines in a phase I trial in melanoma patients using the above mentioned approach (Carreno et al., 2015; Ott et al., 2017; Sahin et al., 2017). Several clinical trials (NCT02950766, NCT02897765) are ongoing to evaluate whether neoantigen cancer vaccines can be combined with CPB therapy (e.g. anti-PD-1 or CTLA4) given its potential to effectively expand the T cell repertoire (Cha et al., 2014;

Robert et al., 2014). In addition, neoantigen-based personalized cancer vaccines are being tested in clinical trial in other cancers such as glioblastoma, breast, pancreatic and colorectal cancers (Türeci et al., 2016).

Whether a similar approach is feasible for MPN patients has not yet been evaluated. MPN diseases are considered hematologic cancers with a relatively low mutational burden. Therefore, it is often assumed that these cancers tend to form less immunogenic neoantigens (Schumacher and Schreiber, 2015). However, many studies rely only on SNVs for the quantification of neoantigen burden and neglect other sources of neoantigens coming from frameshift, fusions and splicing related defects (**Figure 9**). Mutations leading to shift in the reading frame have been demonstrated to act as potent and specific immunogens (Lennerz et al., 2005; Saeterdal et al., 2001). In addition, spontaneous T cell responses specific for MPN driver mutations *JAK2* and *CALR* have been reported, establishing MPN diseases as potentially promising targets for cancer immunotherapy approaches (Holmström et al., 2017, 2016). Therefore, a systematic approach, taking into consideration as many sources of neoantigens as possible for each patient is warranted for MPN diseases.
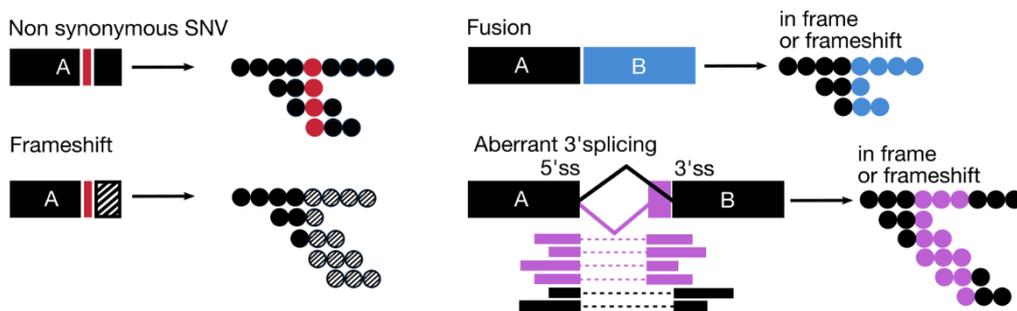
Figure 9: Various classes of mutations can form potential neoantigens. Non-synonymous SNV often leading to protein alterations in a single amino acid. Fusions, frameshift mutations, and splicing alterations can form long stretches of novel peptide sequences.

2 Aim

The main aim of this thesis was to explore novel diagnostic and therapeutic applications using whole transcriptome sequencing data of MPN patients. More specifically, I aimed to

- identify novel makers of clonality in granulocytes of female MPN patients using variant allele frequencies from SNP from the X chromosome,

- establish a fusion detection workflow for systematic identification of fusion genes using transcriptome sequencing data,

- explore the fusion gene landscape of MPN diseases with emphasis on triple-negative MPN and secondary AML patients,

- describe the mutational landscape of the transcriptome including SNVs, Indels in MPN and splicing related defects in PMF patients with mutations in splicing factor gene *SF3B1*,

- determine mis-spliced genes and identify those splicing alterations with putative protein altering consequences in *SF3B1* mutated PMF patients,

- assess the clinical relevance of *SF3B1* mutated MPN patients,

- establish a neoantigen discovery workflow relying solely on transcriptome data,

- explore and mine the neoantigen landscape of MPN patients.

3 RESULTS

The same whole transcriptome dataset of 104 chronic MPN patients and 15 controls were used for results **section 3.1** and **3.2**. Results **section 3.2** also included 9 secondary AML patients. Clinical patient data for results **section 3.1** and **section 3.2** was added to the appendix of this thesis (**Supplementary Table 1**). Column '*upn.clone*' contains unique patient identifiers for results **section 3.1**, column '*unique.patient.identifier*' for results **section 3.2**. In results **section 3.3** targeted sequencing was performed for additional 482 MPN patients.

3.1 Estimation of clonality in granulocytes using variant calls on X-chromosome from RNA-seq data

We took advantage of our transcriptome-wide variant calling to extract a total of 104,722 exonic variants located on the X chromosome. The variants were called and filtered from granulocyte RNA of 104 chronic MPN patients and 15 controls as described in **section 5.1**. In short, variants with a quality by depth of less than 1, covered by less than 10 reads (read depth <10) and located in repetitive regions of the genome or previously reported to reside in genes with known X inactivation escape ("escape genes") were removed. Only SNVs reported as polymorphisms (SNPs) were considered. Finally, 15,934 variants passed all filtering steps.

Variant allele frequencies (VAFs) were extracted for each variant and visualized as histograms (**Figure 10**). The distribution of VAFs had for each sample distinct shapes, ranging from complete unimodal (**Figure 10A**) to bimodal distributions (**Figure 10B, 10C**). The center of the distribution was used to estimate a 'global' ratio of X-inactivation (XCI ratio) for each sample (**Figure 17, 18, and 19 in Material and Methods**). The global XCI ratio represents the extent of granulocyte clonality for each sample.

High concordance of XCI ratio was observed between technical replicates of 4 control samples (**Figure 11**).
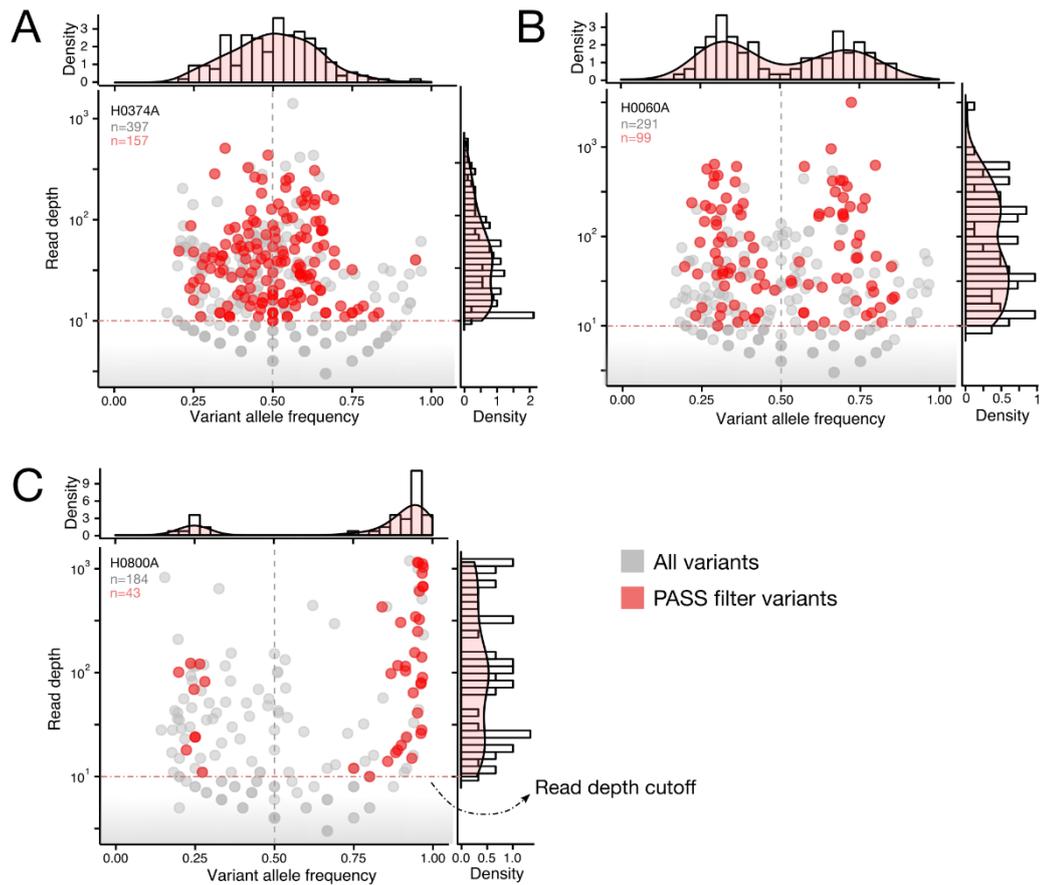
Figure 10: Variant allele frequency (VAF) distributions for three representative MPN patients. VAFs versus the read depth (coverage) for each variant are shown for 3 samples. Histograms and density plots visualize the distribution of VAFs and read depth of these variants. Distribution of VAFs range from unimodal **(A)** to bimodal **(B** and **C)**. The center of the distribution is translated to a global XCI ratio for each sample (see material and methods). The XCI ratio represents the extent of granulocyte clonality (from fully polyclonal (XCI ratio 50:50) to clonal (XCI ratio 100:0) for each sample (XCI ratio=50:50 for **Figure 10A**; XCI ratio=70:30 for **Figure 10B**; XCI ratio=90:10 for **Figure 10C**). The read depth cutoff indicates the threshold for filtering variants covered by less than 10 reads. Only variants colored in red are used for XCI ratio estimation.
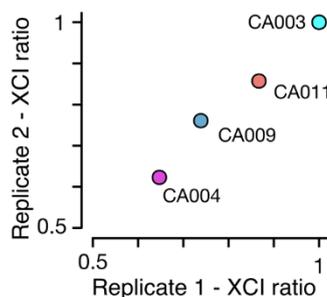
Figure 11: Comparison of XCI ratios for four technical replicates. CA003 is a male, CA0011, CA009, and CA004 are female samples.

First, we tested whether individual genes used for conventional determination of X chromosome inactivation on RNA level like *G6PD*, *MPP1/P55*, *IDS* (Briere and El-Kassar, 1998; el-Kassar et al., 1997) and the master regulator of X-inactivation *XIST (local approach)*, behave similar to the global estimation of XCI ratios (pooling variants from several genes on the X chromosome or *global approach*). For that purpose, we correlated the VAFs of each gene and sample with the XCI ratio estimated using a global approach (global XCI ratio) and calculated a correlation coefficient τ (tau) (**section 3.1.1, Figure 15**). For each gene X, the correlation coefficient τ was plotted against the number of informative (or heterozygous) samples for gene X (**Figure 12**).
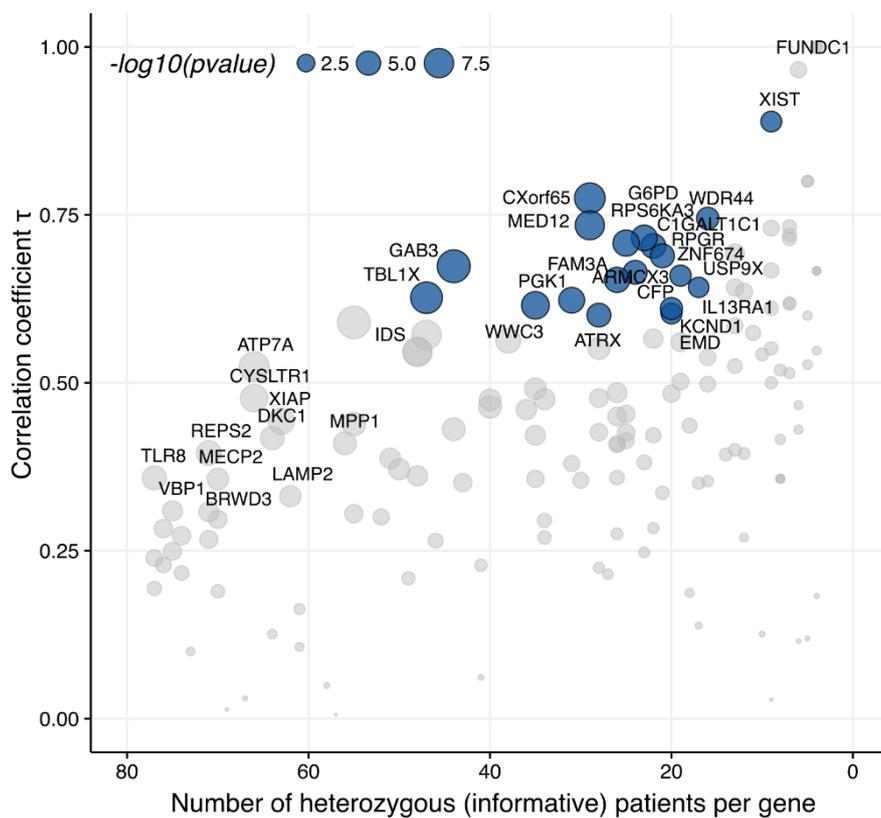
Figure 12: Known and novel marker genes of clonality. For each gene the variant allele frequency was correlated against the global XCI ratio estimate for each sample and a correlation coefficient τ (tau) was assigned to each correlation (**Figure 15**). A non-parametric rank correlation analysis was used to estimate the correlation coefficient τ (or Kendall's τ). The number of heterozygous or informative samples was extracted for each gene and was plotted against the correlation coefficient τ. The size of the circles for each gene, reflects the significance of the correlation -log10(*P-value*). Significant correlations (τ>0.6) were colored in blue (*P-value* <0.05). Selected gene names of 33 genes are shown to allow for better visibility.

Interestingly, the gene with the highest correlation and a significant *P-value* (τ=0.89, *P-value*=0.00024) was *XIST*, a major effector of the X inactivation process (Augui et al., 2011). Other genes with significant correlations (τ>0.6) were *CXorf65, WDR44, MED12,*

G6PD, RPS6KA3, C1GALT1C1, RPGR, GAB3, ARMCX3, ZNF674, FAM3A, USP9X, TBL1X, and *PGK1* (**Figure 12 (highlighted in blue); section 3.1.1, Figure 15**). Among these, *G6PD* is a known gene used for protein or mRNA clonality assays (Prchal and Guan, 1993). Other established markers of clonality such as *IDS* and *MPP1* (Chen and Prchal, 2007; Gregg et al., 2000) were not among the highly ranked genes and had low correlations ($\tau<0.6$). We report new marker genes such as *CXorf65*, *WDR44*, and *MED1*, which have not previously been described.

Next, we analyzed our healthy cohort (7 females and 8 males). Healthy females had an approximate 2:1 ratio of heterozygous versus homozygous variants whereas healthy males with only one X chromosome had exclusively homozygous variants (**Figure 13**).
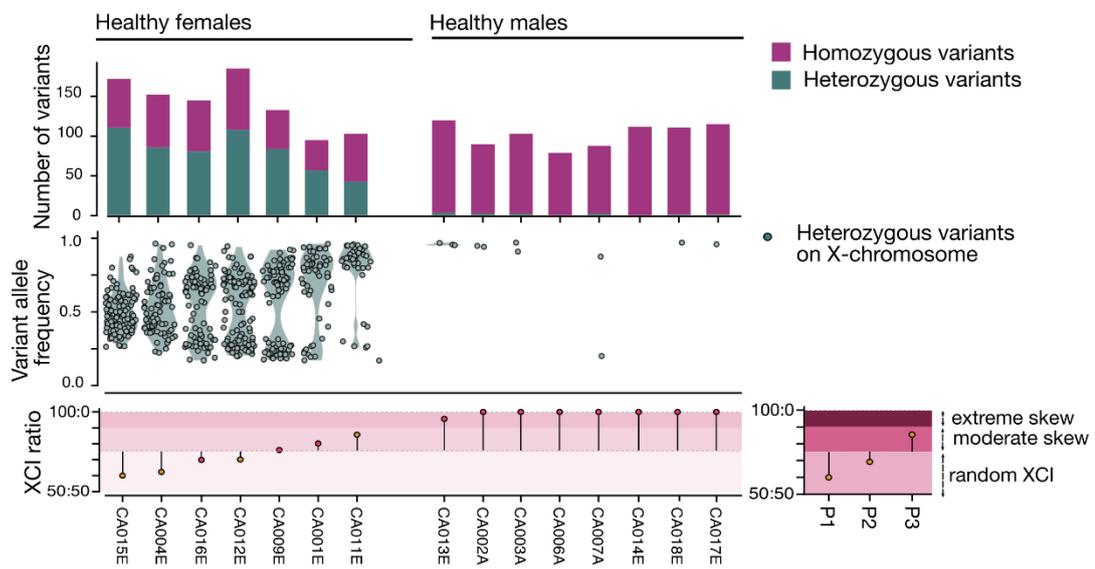
Figure 13: XCI ratios in healthy individuals. (Upper panel) The number of homozygous and heterozygous variants extracted from the X chromosome for each sample are shown as bar graphs. **(Middle panel)** Only heterozygous variants are visualized as violin plots to emphasize the distinct shape of distributions. **(Lower panel)** Corresponding XCI ratios are depicted as needle plots were the mid line is the threshold between random XCI and a moderate skew.

Five females had a random XCI (1/5 close to a moderate skew) and two females had a moderate skew (**Figure 13**). Previous analysis of X-inactivation in unaffected females have shown that the XCI ratio is normally distributed (**section 1.2.2 Figure 4**) (Amos-Landgraf et al., 2006). Given the low number of females in our controls, we could not determine whether the reported XCI ratios in our healthy females reflect a normal distribution as expected in the general population.

Following our healthy individuals, we wanted to assess associations between granulocyte clonality and genetic markers of clonality in female MPN patients such as somatic *JAK2*

and *CALR* mutations. PV patients were highly clonal (median XCI ratio=92:8; n=9) compared to healthy females (median XCI ratio=69:31; n=8) (**Figure 14A**). ET patients with a genetic marker (*JAK2* or *CALR* mutated) were highly clonal (XCI ratio >90:10) in 8/20 patients. Exceptions were present in patient with low burden *JAK2*-V617F mutations or in some patients with aberrations on chromosome 9 (**Figure 14B**). Interestingly, triple-negative ET patients (**Figure 14B**) have a polyclonal pattern of XCI ratio (median XCI ratio=61:39; n=3) comparable to healthy females. The majority of PMF patients were clonal (19/29 with an XCI ratio >90:10) (**section 3.1.1, Figure 16**).
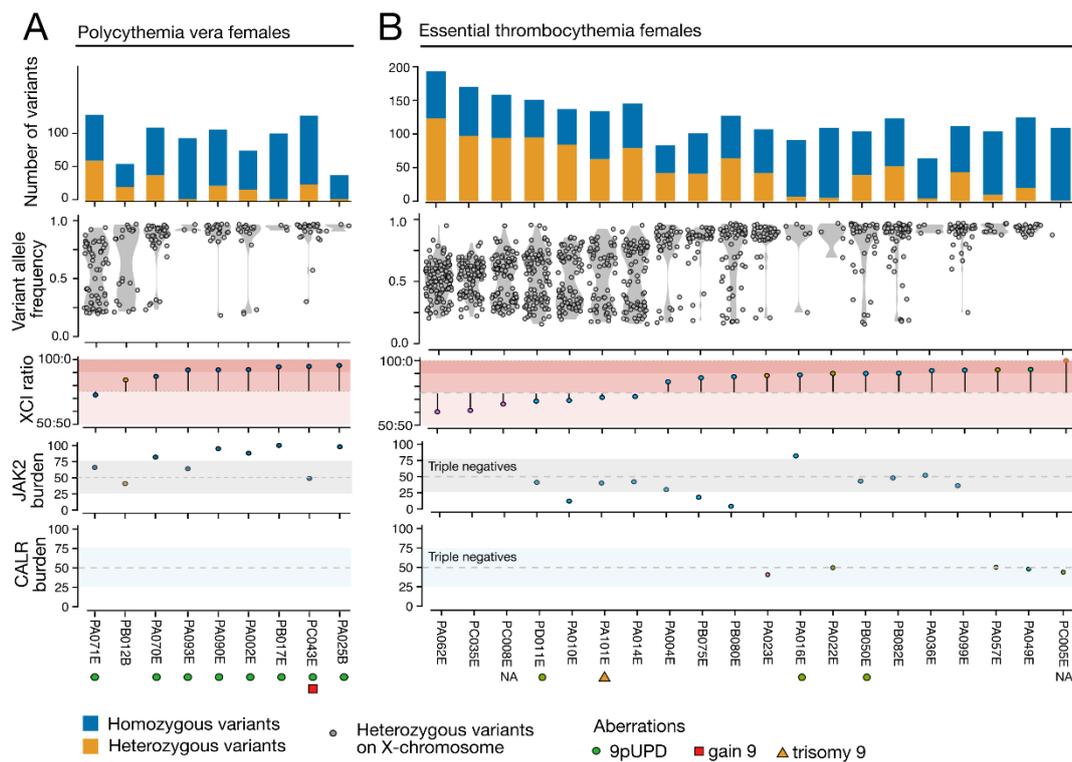
Figure 14: XCI ratios for female PV and ET patients. The first three panels follow the same description as in **Figure 13**. Global XCI ratios were sorted from lowest to highest and plotted accordingly. For each patient *JAK2*-V617F and *CALR* mutational burdens are reported in the two lower panels. Importantly, mutational burdens in *JAK2*-V617F do not directly reflect the proportion of clonal cell populations if aberrations amplifying the *JAK2*-V617F mutation such as 9pUPDs and gains and trisomies of chromosome 9 are present. E.g. a patient with a *JAK2*-V617F burden of 50% and no aberrations has (very likely) a fully clonal cell population (each cell with a heterozygous *JAK2*-V617F mutations). For a patient with a burden of 50% and a clonal 9pUPD aberration the population of clonal cells is calculated as $x=2b/(2+b)$, where b is the *JAK2*-V617F burden [0-1] and $x*100$ is the % of clonal cells.

3.1.1 Supplementary Figures
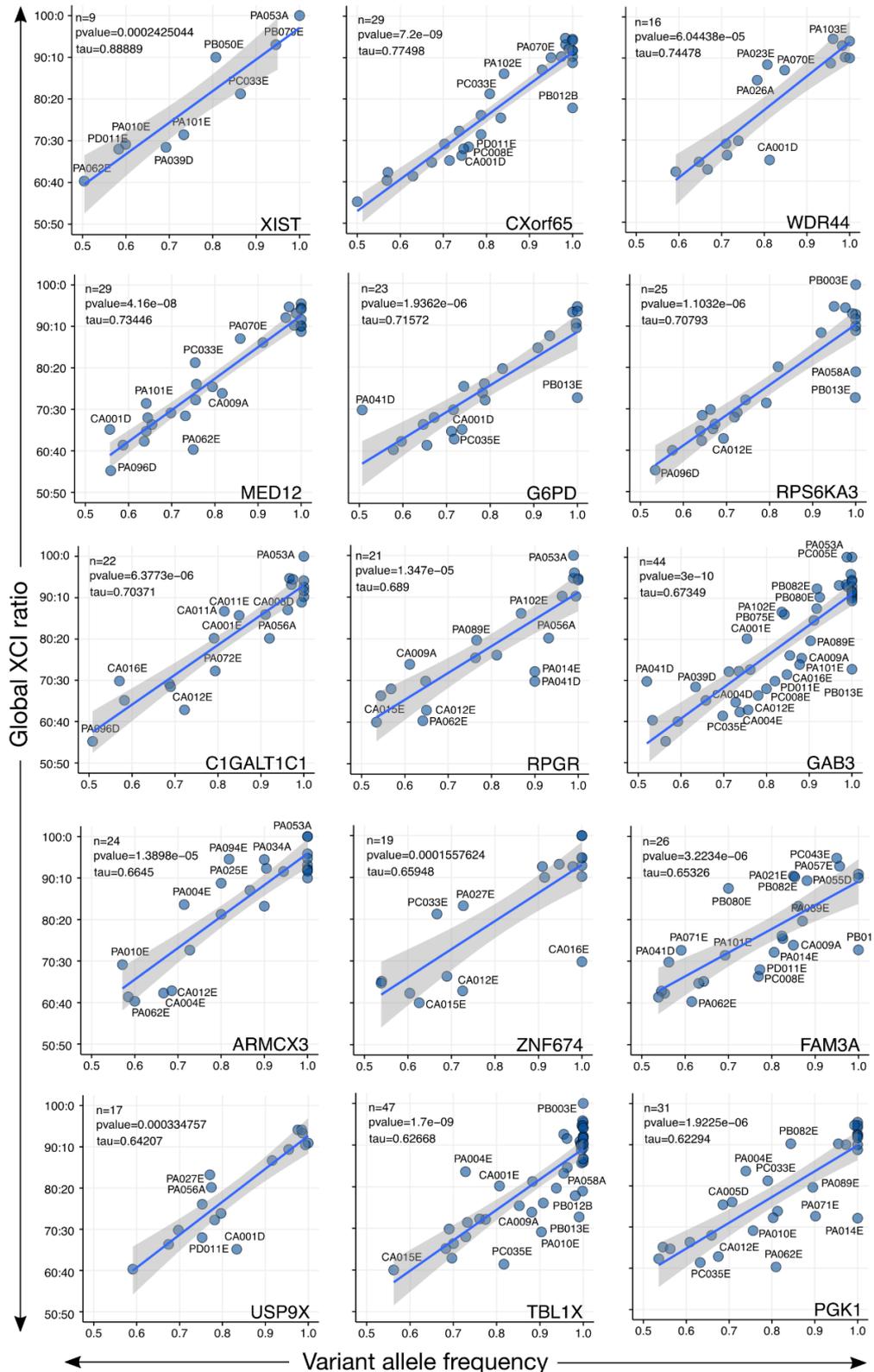
Figure 15: Correlation plots for genes highlighted in Figure 12. SNPs for each gene were extracted. For each patient the variant allele frequency of combined SNPs within a gene (*local approach*) was plotted against the global XCI ratio (*global approach*). A non-parametric rank correlation analysis was used to estimate the correlation coefficient τ (or Kendall's τ).
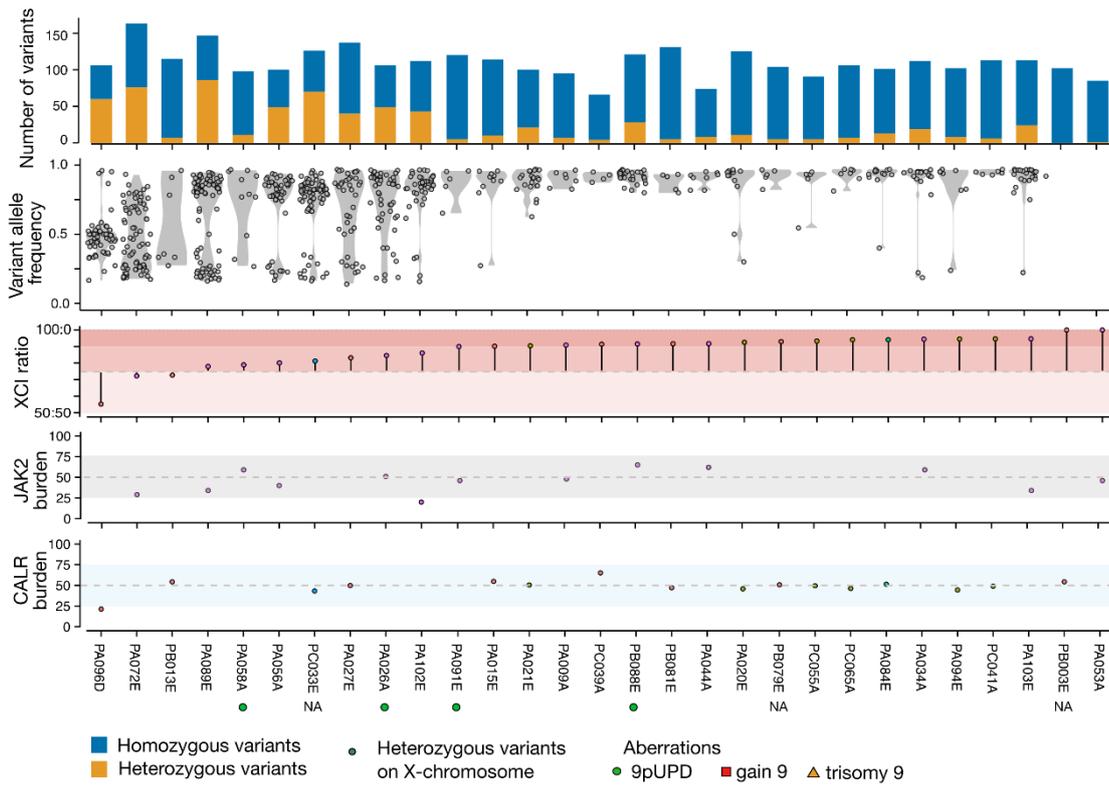
Figure 16: XCI ratios for female PMF patients. Figure description as in **Figure 14**.

3.2 Manuscript #1: Mutational Landscape of the Transcriptome Offers Putative Targets for Immunotherapy of Myeloproliferative Neoplasms

Fiorella Schischlik[1], Roland Jäger[1], Felix Rosebrock[2], Eva Hug[2], Michael Schuster[1], Raimund Holly[1], Elisabeth Fuchs[1], Jelena D. Milosevic Feenstra[1,3], Edith Bogner[1], Bettina Gisslinger[4], Martin Schalling[4], Elisa Rumi[6,7], Daniela Pietra[6], Gottfried Fischer[8], Ingrid Faé[8], Loan Vulliard[1], Jörg Menche[1], Torsten Haferlach [9], Manja Meggendorfer[9], Anna Stengel[9], Christoph Bock[1,10], Mario Cazzola[7], Heinz Gisslinger[4] and Robert Kralovics[1,10]

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria; [2]MyeloPro Diagnostics and Research GmbH, Vienna, Austria; [3]Ludwig Boltzmann Cluster Oncology, Medical University of Vienna, Vienna, Austria; [4]Department of Internal Medicine I, Division of Hematology and Blood Coagulation, Medical University of Vienna, Vienna, Austria; [5]Medical University of Vienna, Vienna, Austria; [6]Unità Operativa di Ematologia, Fondazione IRCCS Policlinico San Matteo, Pavia; [7]Department of Molecular Medicine, University of Pavia, Pavia, Italy; [8]Department for Blood Group Serology and Transfusion Medicine, Medical University of Vienna, Vienna, Austria; [9]MLL Munich Leukemia Laboratory, Munich, Germany; [10]Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria

Running title: Neoantigen discovery in MPN

Scientific category: Myeloid neoplasia

Corresponding author:

Robert Kralovics

CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences

Mail: Lazarettgasse 14, AKH BT25.3, 1090 Vienna

Email: rkralovics@cemm.oeaw.ac.at

Phone: +4314016070027

Fax: +43140160970000

Manuscript: Total word count: 4000, 1 Table, 5 Figures, 43 References

Key Points

- Driver mutations in *CALR* or *MPL* encode for predicted neoantigens that bind MHC class I with high affinity in MPN patients.

- Overall, a substantial portion of MPN patients show evidence of recurrent candidate neoantigens, suggesting a potential use for targeted immunotherapy.

Abstract

Ph-negative myeloproliferative neoplasms (MPNs) are hematological cancers subdivided into entities with distinct clinical features. Somatic mutations in *JAK2, CALR,* and *MPL* have been described as drivers of the disease, together with a variable landscape of non-driver mutations. Despite detailed knowledge of disease mechanisms, targeted therapies effective enough to eliminate MPN cells are still missing. In this study, we aimed to comprehensively characterize in 113 MPN patients the mutational landscape of the granulocyte transcriptome using RNA-seq data and subsequently examine the applicability of immunotherapeutic strategies for MPN patients. Following implementation of customized workflows and data filtering, we identified a total of 13 (12/13 novel) gene fusions, 231 non-synonymous SNVs, and 21 Indels in 106/113 patients. We found a high frequency of *SF3B1* mutated PMF patients (14%) with distinct 3'splicing patterns, many of these with a protein altering potential. Finally, from all mutations detected, we generated a virtual peptide library and used NetMHC to predict 149 unique neoantigens in 62% of MPN patients. Peptides from *CALR* and *MPL* mutations provide a rich source of neoantigens due to their unique property to bind many common MHC class I molecules. Finally, we propose that mutations derived from splicing defects present in *SF3B1* mutated patients may offer an unexplored neoantigen repertoire in MPN. We validated 35 predicted peptides to be strong MHC class I binders through direct binding of predicted peptides to MHC proteins *in vitro*. Our results may serve as a resource for personalized vaccine or adoptive cell-based therapy development.

Introduction

Ph-negative myeloproliferative neoplasms (MPN) are clonal hematological malignancies with frequent mutations in *JAK2*, *CALR* and *MPL* gene. Transformation to blast phase or secondary acute myeloid leukemia (post-MPN sAML) can occur. MPN patients testing negative for disease driver gene mutations are often referred to as "triple-negative"[1,2]. Targeted and whole-exome sequencing effort have identified single nucleotide variants (SNVs) and small insertions and deletions (Indels) with low overall mutation frequency in essential thrombocythemia (ET), polycythemia vera (PV), and primary myelofibrosis (PMF)[1,3,4]. Other mutation classes, such as gene fusions and splicing-related defects, have never been systematically mapped and explored in MPN. RNA-seq is the preferred method for the transcriptome wide discovery of these mutation classes[8]. One of the aims of our study was to extensively characterize the mutational landscape of MPN patients using a transcriptome-based approach.

Spontaneous T cell responses of CD8[+] and CD4[+] T lymphocytes against MPN driver mutations *JAK2*[9] and *CALR*[10,11] have been described, establishing MPN as potentially immunogenic neoplasms and suggesting the usage of personalized cancer vaccines, checkpoint inhibitor, adoptive T-cell therapy or combinations of those[12] for the treatment of MPN. Neoantigens that also function as oncogenic driver mutations in the respective disease are highly attractive targets since they are shared between patients, essential for tumor survival and expressed in the disease causing clone[13–15]. However, the huge diversity in HLA haplotypes significantly restricts the abundance of neoantigens. Therefore, further exploration of the neoantigen repertoire in MPN is essential in order to evaluate if the MPN somatic mutational landscape offers the possibility to identify candidate neoantigens for immunotherapy. In this study, we provide a framework for further systematic identification of MPN neoantigens for mutation classes such as SNVs, Indels, gene fusions, and splicing abnormalities[16], solely based on RNA-seq data obtained from granulocytes – the most abundant clonal myeloid cell type in MPN. Many of these mutation classes lead to frameshifts and introduction of novel amino acid sequences and therefore the distinction from non-mutated protein sequence is more dramatic (compared to SNVs). These mutation classes have also been shown to act as potent and specific immunogens[17,18]. We predicted neoantigens for each patient in our MPN cohort on a personalized level taking into account each patient's MHC class I (MHCI) genotype. Finally, we validated the peptide:MHCI binding affinity for a subset of candidate neoantigenic peptides and thereby established a resource for immunotherapy in MPN.

Material and Methods

Data access

Whole-transcriptome sequence data has been deposited at the European Genome-phenome Archive, under accession number **EGAS00001003486**.

Code availability

The source code for the implementation of the workflows described in this study is available for download and use on GitHub (http://github.com/sp00nman/rnaseqlib).

Patient and control samples

A total of 104 MPN patients and 15 healthy controls were recruited into the study at the Medical University of Vienna, Austria and 9 post-MPN sAML patients from the Fondazione IRCCS Policlinico San Matteo Pavia, Italy. (**Figure 1A, Table 1**). The study was reviewed and approved by local ethics committees and all patients and healthy controls signed an informed consent in accordance with the Declaration of Helsinki. Criteria applied for the diagnosis of patients were described previously[1] (**Supplemental Table 1**).

Transcriptome library preparation and sequencing

Total RNA was isolated applying standard procedures from peripheral blood granulocytes. Next-generation sequencing of cDNA libraries of polyA-enriched RNA were generated using different library preparation kits and sequenced in various configurations on an Illumina HiSeq2000 instrument as listed in **Supplemental Table 2**.

Driver mutation analysis

The presence of mutations and mutational burden of *JAK2*, *MPL* and *CALR* were determined in all patients as described in Klampfl et al., 2013[1].

Transcriptome data analysis for fusion detection and validation

Sequencing reads were mapped and processed independently with fusion detection tools *defuse*, *tophatfusion* and *soapfuse* using human genome version GRCh37 or hg19. The results of these fusion detection algorithms were merged and further filtered as described in Supplemental Methods. The fusion candidates that passed the filtering steps were validated by reverse transcription polymerase chain reaction (RT-PCR) using RNA isolated from human peripheral granulocytes, followed by Sanger sequencing of the PCR products as described in Supplemental Methods.

Variant calling on RNA sequencing data and targeted re-sequencing of patients' genomic DNA

For the purpose of identification of single nucleotide variants (SNVs), small insertions and deletions (Indels), we developed a workflow combining GATK's (version 3.4-46) [19] "*Best Practice workflow for RNA-seq*" with in-house algorithms for reducing false-positive calls (see Supplemental Methods). Variants were filtered to enrich for somatic mutations as described in Supplemental Methods. For 77 patients additional targeted re-sequencing of genomic DNA was performed (Supplemental Methods).

Estimating expression values on a gene level and differential expression analysis

Sequencing reads were trimmed (Trimmomatic) and aligned with the STAR aligner to the UCSC hg38 reference genome. Reads overlapping transcript features were counted with the *summarizeOverlaps* function of Bioconductor library *GenomicAlignments*. A detailed description of the workflow is described in Supplemental Methods. The Bioconductor package *DESeq2* was then used to model the data set and call differentially expressed genes.

Identification of aberrant 3' splicing in SF3B1-mutated patients and validation

RNA-seq reads were aligned to the human genome as described in the *Variant calling on RNA sequencing data* section. Splice junction coverage files *SJ.out.tab* from STAR aligner were used for further splice junction analysis. Aberrant 3' splicing was detected following published instructions[20]. Novel splice junctions were defined as described in Supplemental Methods. Differential splice site usage was tested with the *testForDEU*

function of the *DEXSeq* Bioconductor package (version 1.14.2)). Obtained *P-values* were adjusted for false discovery rate (FDR) using the Benjamini-Hochberg procedure. Hierarchical clustering was performed using the R package *pheatmap*. Percent-spliced-in (PSI) scores for alternative 3' splice sites (3'ss) were calculated as described in **Figure 4A**. Aberrant 3'ss were validated with fragment length analysis using capillary electrophoresis as described in Supplemental Methods.

Neoepitope prediction and prioritization

We generated *FASTA* sequences of peptides from mutation classes such as SNVs, Indels, fusions and aberrantly spiced genes as described in Supplemental Methods. Predicted HLA-I alleles for each patient (*seq2HLA*[21], release 2.2) and the mutated peptide sequences were used as input for NetMHCpan[22] (version 4.0) to predict binding affinities of 8, 9 and 10-mer peptides to MHCI proteins. We further filtered the NetMHCpan results based on criteria defined in Supplemental Methods.

Peptide:MHCI binding validation through peptide exchange technology

HLA Flex-T monomers (HLA*A03:01, HLA*A11:01, HLA*B07:02, HLA*B08:01) as well as positive and negative control peptides with known binding affinities were purchased from BioLegend. The peptides were custom-designed by Eurogentec. Monomers were mixed 1:1 with the appropriate peptides and incubated for 30min under UV light (365nm) and afterwards kept for 30min in the dark. 1μl of this mixture was then used 1:1000 in duplicates in the ELISA following the Flex-T HLA Class I ELISA protocol from BioLegend. The read-out is shown as mean absorbance.

Statistical analysis

Pairwise associations of genes and diseases were calculated using Fisher's exact test followed by FDR correction (Benjamini-Hochberg). All statistical analysis was performed using R (version 3.2.3 2015-12-10).

Results

Using whole-transcriptome sequencing, we analyzed RNA from granulocytes of 104 patients with chronic MPN (32 ET, 17 PV, 55 PMF) and 9 patients that transformed to post-MPN sAML, as well as 15 healthy controls (n=128). (**Figure 1A, Table 1, Supplemental Table 1, 2**). We established a transcriptome-based bioinformatics workflow to enable the identification of gene fusions, variants (SNVs & Indels) and splicing abnormalities (**Figure 1B, Supplemental Figure 1**).

Fusion discovery

To establish the fusion discovery workflow, we first performed RNA sequencing of 20 patient samples from various hematological malignancies with known fusion genes validated with fluorescence *in situ* hybridization. Except for fusion rearrangements involving promoter exchange (most *IGH*-fusions), which are known as difficult to capture with RNA-seq, we were able to recover all validated fusions and identify novel fusion partners (**Supplemental Figure 2A, 2B and Supplemental Table 3**). Next, we applied the same workflow to the MPN patient cohort and identified a total of 13 fusions that passed all filtering criteria (**Supplemental Figure 3A–3F, Supplemental Table 4**). All cDNA breakpoints were validated by Sanger sequencing (**Supplemental Figure 4**). Eleven out of 13 fusions were detected in chronic-phase MPN and 2 in post-MPN sAML patients (**Figure 1C, Supplemental Figure 5A**). The majority (8/13) of fusions had fusion partners *in trans* (i.e. on different chromosomes) and were, therefore, suspected to be caused by translocations. The remaining 5 fusions were formed *in cis* by inversions (n=3), by a large-scale deletion (n=1) and by a tandem duplication (n=1) (**Supplemental Figure 5A, 5B**).

Fusion discovery results were merged with cytogenetic aberrations data obtained from Affymetrix SNP 6.0 arrays to verify if breakpoints coincided (**Supplemental Figure 5B and Supplemental Table 4, 5**). A large chromosome 13q deletion detected by SNP arrays in a PMF patient lead to a *FRY-MYCBP2* gene fusion (**Supplemental Figure 4D**). None of the other fusion genes overlap with previously detected chromosomal aberrations or were not within the detection range of SNP arrays. RNA expression values for most fusion genes were low (mean=3.2, SD=2.3 SRPKM), except for *FRY-MYCBP2* (106.7 SRPKM), (**Supplemental Figure 5C**), however, the respective wildtype gene expression

value for each gene forming part of the fusion varied considerably (**Supplemental Figure 5D**).

Among the 13 fusions found in MPN and post-MPN sAML patients, *BCR-ABL1* t(8;22) and *INO80D-GPR1* inv(2) were the only *in frame* fusions. (**Supplemental Figure 4A, 4H**). For the *INO80D-GPR1* inversion, we identified the genomic breakpoint region to be in intron 8 (genomic position chr2:206,584,716; hg19) of *INO80D* and exon 4 (chr2:206,750,070; hg19) of *GPR1* (**Supplemental Figure 4I**). We screened additional 96 MPN patients for presence of *INO80D-GPR1*, however no other occurrence was found, suggesting that this inversion is likely a private somatic aberration.

Variant calling at transcriptome-level and mutation frequencies in MPN

Next, we applied the variant discovery workflow on RNA data and called variants in 87 genes recurrently mutated in myeloid malignancies and/or MPN (**Supplemental Table 7**). After applying RNA-specific filtering criteria (**Supplemental Methods - RNA-specific filter a-e**), we identified 262 variants in 59/87 genes in 106/113 patients. Of the 262 variants, 221 were non-synonymous SNVs, 13 deletions, 18 insertions and 10 stop-gain mutations (**Figure 1C, Supplemental Table 8**).

To cross-validate variant calling based on RNA, 77 matching genomic DNA samples from the same patient were sequenced via the TruSight Myeloid targeted re-sequencing panel (Illumina) (**Supplemental Table 9**). We found an overall concordance of 82.2% for SNVs (10x coverage, VAF>0.1), which is consistent with published findings[23,24] (**Supplemental Figure 6D and Supplemental Table 10**). To test whether variant allele frequencies (VAF) vary significantly between RNA and DNA, we compared VAFs for *JAK2*-V617F assessed with allele-specific PCR on genomic DNA with VAFs obtained from RNA sequencing. We observed a highly significant correlation between these two methods (R^2=0.86, *P-value*=2.2x10^{-16}) (**Supplemental Figure 6C**). In addition, most of the called RNA variants were called on clonal cell populations and were expressed in myeloid cells. More than half of the variants were annotated in the COSMIC database as 'confirmed somatic mutation' and were frequently located in oncogenic hotspots (**Figure 2, Supplemental Table 11 &12, and Supplemental Figure 7**).

We identified a high frequency of mutations in epigenetic modifiers such as *TET2* (18.8% ET, 21.8% PMF, 17.7% PV) and *DNMT3A* (3.2% ET, 12.7% PMF, 11.8% PV), as well as an unexpectedly low frequency of *ASXL1* mutations in PMF (3.13% ET, 3.64% PMF,

11.76% PV) compared to published data[3,4] (**Figure 2, Supplemental Table 11**). We also found unexpectedly high frequencies of mutations in *NOTCH1* (3.1% ET, 7.3% PMF, 17.7% PV), *SH2B2* (3.1% ET, 3.6% PMF, 17.7% PV) and *SF3B1* (0% ET, 14.6% PMF, 0% PV).

SF3B1-mutated patients display a distinct pattern of 3' splicing abnormalities

The high frequency of mutations in the S*F3B1* gene in our RNA cohort was restricted to PMF patients (9/55; 16.3%). Furthermore, *SF3B1* mutations were mutually exclusive with mutations in other genes of the splicing machinery (*U2AF1, SRSF2,* and *SF3A1*) and the patients were either *JAK2*-V617F or *CALR* mutation positive (**Figure 3A**). The 9 PMF patients with *SF3B1* mutations had non-synonymous amino acid changes (K700E, K666N/R/T and R594L) and except for R594L, all mutations coincided within mutation hotspot sites when compared to mutation data extracted from the COSMIC database[25] (**Figure 3B**). To evaluate if these mutations introduce splicing abnormalities, we performed differential junction expression analysis comparing PMF patients carrying canonical *SF3B1* mutations (K666/N/R/T, and K700E) with the remaining 47 *SF3B1* wildtype PMF patients and 15 healthy controls. Using *DEXseq* we identified a total of 330,031 junctions of which a total of 850 differentially expressed junctions (adjusted *P-value*<0.1, log_2(fold change-FC)>0) exhibited a novel 3' or 5' splice site in the *SF3B1*-mutant patients (**Supplemental Table 13**). Aberrant novel 3' splice sites (3'ss) were around 4.9-fold more abundant than 5'ss (705 vs 145; **Figure 3C**). Focusing on aberrant 3'ss, we calculated the nucleotide distance for the aberrant 3'ss from its canonical splice site and observed a high enrichment between nucleotides 10 to 30 (**Figure 3D**). For 271 aberrant 3'ss within 10-30 nucleotides from its canonical splice site we could show distinct clustering of canonical (K700E, K666N/R/T) *SF3B1* mutations and non-mutated or non-canonical *SF3B1* (R594L) mutations (**Figure 3E, Supplemental Table 14**). We observed both shared and mutation-specific aberrant 3' splicing patterns between K700E and K666N/R/T (**Figure 3F**). In order to examine if other mutations in the splicing machinery (*U2AF1, SRSF2,* and *SF3A1*) have a similar impact on 3'ss, we included patients with mutations in these genes and performed hierarchical clustering. As previously reported for other cancer types[26], only K700E and K666N/R/T *SF3B1* mutations showed a strong penetrance for the 3' splicing abnormality phenotype (**Supplemental Figure 8**). To examine if the 3' splicing abnormalities we identified in PMF are shared among other malignancies, we compared differentially spliced junctions

among 6 additional cancer types where mutations in the splicing machinery have been previously described[27–30]. This analysis revealed that the 3' splicing abnormalities we discovered in *SF3B1*-mutated PMF are disease-specific. However, genes with the highest percentages of aberrant transcripts showed more frequent overlap among the analyzed cancer types (**Supplemental Figure 9**). When we quantified the overlap across all cancer types, 2 genes with aberrant splicing (*OXA1L* and *SLC3A2*) were shared in all 7 analyzed cancer types (**Supplemental Figure 9C**). In *SF3B1*-mutated PMF patients we detected 141 unique splicing aberrations (**Supplemental Table 15**).

Identification of 3' splice sites with highest impact on predicted amino acid changes

Next, we aimed to identify those 3' splice junction sites with the greatest impact on the protein sequence. We calculated the *percent spliced in* (PSI) value for each splice site and patient (**Figure 4A**), which represents the percentage of the abnormal transcript of the total number of transcripts. A total of 43 novel 3' splice junctions showed a median PSI ≥ 20% for *SF3B1*-mutated patients (**Supplemental Table 16**). After predicting the impact of each splicing abnormality on the amino acid sequence, we selected 20 splicing abnormalities that resulted in a frameshift and truncation of the protein or introduced a stretch of novel amino acids *in frame* to the native protein sequence. Hierarchical clustering of PSI values of aberrant 3'ss in relation to canonical splice sites showed preferential usage of aberrant 3'ss by *SF3B1*-mutated patients compared to *SF3B1*-wildtype patients (**Figure 4B**). Aberrant 3'ss usage was also dependent on specific *SF3B1* mutants (K700 and K666) (**Figure 4C**). We confirmed the presence of 19/20 abnormal splice events detected by RNA-seq using an alternative methodology (fragment length analysis of RT-PCR products) in 4 patients (P009A, P040B, P048A, P069A) (**Figure 4D, 4E**). In **Figure 4F** we summarized several attributes for the selected 20 genes. Interestingly, of the 16 *out of frame* truncations, only the *TTI1* gene showed a significant (*log2*(FC)=-0.38; adjusted *P-value*=0.034) decrease of the mRNA level, suggesting that non-sense mediated decay (NMD) does not influence the transcript abundance in most of the selected genes. Moreover, *OXA1L* showed even a significant elevation (*log2*(FC)=0.86; adjusted *P-value* =2.02x10^{-7}) of the mRNA level despite the presence of a truncation (**Supplemental Table 17**).

RNA-based workflow for systematic identification of neoantigens

Published workflows for neoantigen discovery rely on exome sequencing combined with transcriptome sequencing to focus only on expressed variants[31–33]. In our approach, we bypassed the exome sequencing step by calling the expressed mutation classes directly from transcriptome data. In addition, we called MHCI haplotypes directly from transcriptome data for 113 MPN patients and controls (**Supplemental Table 18**). For each of the *HLA-A*, *HLA-B*, and *HLA-C* genes, we identified the 4 most frequent alleles in our cohort (**Supplemental Table 19**). Next, we used the peptide-to-MHCI (peptide:MHCI complex) binding prediction algorithm NetMHCpan[22] to predict tumor specific neoantigens for MHCI presentation.

To examine the number of predicted neoantigens for the 12 most common HLA alleles in our patient cohort, we generated a virtual peptide library based on the mutations detected at RNA level. This peptide library consisted of 541 patient-specific (or 149 unique) peptides predicted to have a binding affinity to one of the 12 common MHCI proteins. On cohort level, 102 were predicted to have a strong binding (%Rank <0.5), while 439 were weak binders (%Rank <2) (**Figure 5A, Supplemental Table 20**). In 86% of peptide:MHCI pairs, the peptide was derived from frameshift causing mutations emphasizing their potential role as a source of neoantigens (**Supplemental Figure 10**). Although *CALR* occupied the majority (n=403; 86.9%) of peptide:MHCI pairs, neoantigens derived from frameshift mutations in *TET2* (n=16; 3.5%) were also present, but none of these were shared within the MPN cohort.

MPN driver mutations CALR and MPL are a rich source of neoantigens

Driver mutations in MPN are well-characterized and are represented by *JAK2*-V617F, *MPL*-W515K/L/A and frameshift mutation in *CALR*. Since the more frequent *MPL*-W515L and *MPL*-W515K point mutation were not present in our cohort, we tested these separately for peptide:MHCI interactions. Predictions of 4 of the most common *HLA-A*, *HLA-B,* and *HLA-C* alleles resulted in 42 peptide:MHCI pairs for *CALR* and 17 peptide:MHCI pairs for *MPL*-W515K/L/A (**Figure 5B, Supplemental Table 20, 21**). The *CALR* ins5 mutation form more unique predicted neoantigens than the *CALR* del52 mutation towards the N-terminus of the *CALR* protein (**Figure 5E**). We did not identify any binding peptides derived from the *JAK2*-V617F driver mutation for any of the HLA genotypes tested.

Identification of neoepitopes in aberrantly spliced genes

Next, we wanted to examine if the peptide sequences derived from abnormally spliced genes in *SF3B1*-mutated patients may serve as putative neoantigens. Out of the 850 differentially expressed junctions, 21 aberrant splice sites (**Supplementary Table 13**) were not expressed in *SF3B1*-wildtype patients. While these 21 splice sites may, from a tumor-specificity point of view, present suitable candidates for neoantigens derived from aberrant splice sites, none of these splice sites passed other equally important criteria defined in Supplemental Methods. Therefore, we focused our analysis on our previously defined set of 19 genes affected by aberrant splicing listed in **Figure 4F**, leaving the question of true tumor specificity open. We used NetMHCpan to predict peptide:MHCI binding applying the same criteria for weak and strong binders as previously described. 169 peptides had at least 1 MHCI protein with a predicted binding %Rank <2 (**Figure 5C, Supplemental Table 22**). We counted the number of predicted neoantigens considering each patient's personal MHCI haplotype. Patients with *SF3B1* and *CALR* co-occurring mutations had the highest number of neoantigens presented by patient's own MHCI variants (68.0 average number of predicted weak binders and 21.6 predicted strong binders) (**Figure 5D, Supplemental Table 23**).

In vitro-validation of predicted peptides to MHCI protein

To validate the predicted peptide:MHCI binding, an ELISA-based peptide exchange assay was used[34]. We experimentally tested 35 peptides predicted to bind either weakly (n=15) or strongly (n=20) to 4 commercially available HLA monomers (A*03:01, A*11:01, B*07:02, and B*08:01). Control peptides of known strong or weak binding affinity were available for A*03:01, A*11:01, and B*07:02 (**Figure 5F, Supplemental Table 24**). We defined a threshold for binders versus non-binders as 2 standard deviations from the mean of the absorbance values of 3 positive control peptides in the ELISA assay. Using an *in vitro* binding assay, we could validate 23 of the 35 (65.7%) predicted peptide:MHCI interactions.

Discussion

We studied the global transcriptome of 113 MPN patients using RNA-seq technology and described in detail their mutational landscapes. We established an RNA-seq based pipeline utilizing high quality transcriptome data as an alternative method for neoantigen discovery. This is the first study that globally annotated the MPN transcriptome with the aim to identify and systematically mine MPN-relevant, putative neoantigens for targeted immunotherapy of MPN.

Another novel aspect of this study is the global analysis of gene fusions in MPN that are missed by targeted or exome sequencing. We detected and annotated gene fusions in a systematic and unbiased high-throughput transcriptome-wide approach. Fusion genes found in chronic MPN (*JAK2* or *CALR*-positive) were scarce. Except for known and *in frame BCR-ABL1* oncogenic fusion found in one PMF patients, 10 fusions detected in chronic MPN were novel and *out of frame*. For some fusions we observe high expression of wildtype gene and low expression of the fusion transcript, which is suggestive for fusion formation through trans-splicing events as suggested by previous studies[35,36]. Lack of recurrence for any of the identified fusions prohibited any mechanistic studies. Two out of 9 post-MPN sAML patients had fusion genes (*INO80D-GPR1* and *XPO5-RUNX2*) not previously described. Surprisingly, no typical *de novo* AML fusion oncogenes were detected, emphasizing the distinct nature of disease progression in post-MPN sAML compared to *de novo* AML[37]. Interestingly, predicted protein products of fusion genes, in particular frameshifts or their junction regions, were not a significant source of neoantigens.

We focus our analysis on 86 genes frequently mutated in clonal myeloid diseases. In addition, we reported HLA genotypes extracted from RNA-seq data in a cohort of 113 MPN patients. This enabled us to make a truly personalized prediction of neoantigen occurrence for each MPN patient. Although the number of Indels and SNV identified per patient was low as expected for MPN [4], we could still identify 149 unique neoantigens in 62% of MPN patients that could have potential use for personalized cancer immunotherapy approaches.

Having defined a gene set with the highest impact on protein sequence, we hypothesized that splicing abnormalities induced by canonical *SF3B1* mutations may serve as potential neoantigens for MPN. Several findings strengthened our assumptions that frameshift mutations caused by aberrant splicing might indeed serve as potent neoantigens. We

could show high expression and high VAF of the mutant copy of *SF3B1*. The aberrantly spliced genes we described did not show frequent down-regulation on mRNA levels (with exception of *TTI1*), arguing against the involvement of nonsense-mediated decay of aberrant transcripts as previously suggested by other studies[29,38]. Furthermore, for 15 peptides derived from abnormally spliced genes we showed strong affinity to common MHCI molecules using an *in vitro* binding assay. Although additional studies are needed to demonstrate their immunogenicity, peptides derived from splicing defects may represent recurrent immunotherapy targets, perhaps as valuable as peptides derived from the frameshifted *CALR* oncogenic MPN driver. Finally, although our research did not provide evidence of these aberrant transcripts on a protein level, a recent study identified so called 'neojunctions' in *SF3B1* mutated cancers and could confirm presence of these neojunctions on protein level using mass spectrometry approaches[39]. We think, it would be of great interest for future research to formally show that our candidate splicing derived putative neoantigens are indeed translated to proteins and therefore potentially capable of being presented in MHCI complexes on the cell surface.

Another important question is the tumor specificity of the splicing defects. Aberrant splicing at a specific junction may also occur at lower frequency in healthy cells so that targeting the resulting protein sequence may cause autoimmunity and toxicity[40,41]. Our results show for several candidate genes weak presence of the aberrant transcripts in *SF3B1* non-mutated patients. Nevertheless, a search through the human proteome database with the translated novel protein sequences did not yield any significant matches, excluding the possibility that any of these aberrant transcripts are merely alternative isoforms. Thus, further experimental evidence to show tumor specificity on protein level is required to classify whether aberrant 3'ss are true neoantigens or tumor-associated antigens overexpressed in tumor cells.

Our future studies will be aimed at annotating the candidate neoantigenic peptides identified in this study for the ability to induce T cell activation both, in healthy controls and in *SF3B1*-mutated MPN patients to prove that these are truly immunogenic antigens. In summary, our study may serve as a foundation and resource for the development of individualized vaccines or adoptive cell-based therapies for MPN, in particular for PMF patients, where the unmet medical need is greatest.

Acknowledgements

Authorship Contributions

F.S. and R.K. conceived and designed the experiments and wrote the manuscript. F.S, R.K., R.J., F.R., E.H., R.H., J.M., E.B. performed the research and experiments. F.S., R.K., M.S., C.B. analyzed, interpreted the data and performed statistical analysis. H.G., B.G., M.S., M.C., E.R., D.P., G.F., I.F., T.H., M.M., A.S., L.V., J.M. contributed materials and analysis tools. All authors contributed to the final version of the manuscript.

Conflict of Interest Disclosures

Robert Kralovics: Pharma Essentia (honoraria, research funding, consultancy); AOP Orphan Pharmaceuticals (honoraria, consultancy); MyeloPro Diagnostics and Research (stock ownership); Qiagen (consultancy); Novartis (honoraria)

References

1. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N. Engl. J. Med.* 2013;369(25):2379–90.

2. Milosevic Feenstra JD, Nivarthi H, Gisslinger H, et al. Whole-exome sequencing identifies novel MPL and JAK2 mutations in triple-negative myeloproliferative neoplasms. *Blood*. 2016;127(3):325–332.

3. Lundberg P, Karow A, Nienhold R, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood*. 2014;123(14):2220–8.

4. Nangalia J, Massie CE, Baxter EJ, et al. Somatic *CALR* Mutations in Myeloproliferative Neoplasms with Nonmutated *JAK2*. *N. Engl. J. Med.* 2013;369(25):2391–2405.

5. Klampfl T, Harutyunyan A, Berg T, et al. Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. *Blood*. 2011;118(1):167–176.

6. Milosevic JD, Kralovics R. Genetic and epigenetic alterations of myeloproliferative disorders. *Int. J. Hematol.* 2013;97(2):183–97.

7. Bartels S, Lehmann U, Büsche G, et al. SRSF2 and U2AF1 mutations in primary myelofibrosis are associated with JAK2 and MPL but not calreticulin mutation and may independently reoccur after allogeneic stem cell transplantation. *Leukemia*. 2015;29(1):253–5.

8. Cieślik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. *Nat. Rev. Genet.* 2017;

9. Holmström MO, Hjortsø MD, Ahmad SM, et al. The JAK2V617F mutation is a target for specific T cells in the JAK2V617F-positive myeloproliferative neoplasms. *Leukemia*. 2017;31(2):495–498.

10. Holmström MO, Riley CH, Svane IM, Hasselbalch HC, Andersen MH. The CALR exon 9 mutations are shared neoantigens in patients with CALR mutant chronic myeloproliferative neoplasms. *Leukemia*. 2016;30(12):2413–2416.

11. Holmström MO, Martinenaite E, Ahmad SM, et al. The calreticulin (CALR) exon 9 mutations are promising targets for cancer immune therapy. *Leukemia*. 2017;

12. Wirth TC, Kühnel F. Neoantigen Targeting—Dawn of a New Era in Cancer Immunotherapy? *Front. Immunol.* 2017;8:1848.

13. Shono Y, Tanimura H, Iwahashi M, et al. Specific T-cell immunity against Ki-ras peptides in patients with pancreatic and colorectal cancers. *Br. J. Cancer*. 2003;88(4):530–536.

14. Ichiki Y, Takenoyama M, Mizukami M, et al. Simultaneous cellular and humoral immune response against mutated p53 in a patient with lung cancer. *J. Immunol.* 2004;172(8):4844–50.

15. Sharkey MS, Lizée G, Gonzales MI, Patel S, Topalian SL. CD4(+) T-cell recognition of mutated B-RAF in melanoma patients harboring the V599E mutation. *Cancer Res*. 2004;64(5):1595–9.

16. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic *SF3B1* Mutation in

Myelodysplasia with Ring Sideroblasts. *N. Engl. J. Med.* 2011;365(15):1384–1395.

17. Lennerz V, Fatho M, Gentilini C, et al. The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. *Proc. Natl. Acad. Sci.* 2005;102(44):16013–16018.

18. Saeterdal I, Bjorheim J, Lislerud K, et al. Frameshift-mutation-derived peptides as tumor-specific antigens in inherited and spontaneous colorectal cancer. *Proc. Natl. Acad. Sci.* 2001;98(23):13255–13260.

19. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303.

20. DeBoever C, Ghia EM, Shepard PJ, et al. Transcriptome Sequencing Reveals Potential Mechanism of Cryptic 3' Splice Site Selection in SF3B1-mutated Cancers. *PLOS Comput. Biol.* 2015;11(3):e1004105.

21. Boegel S, Löwer M, Schäfer M, et al. HLA typing from RNA-Seq sequence reads. *Genome Med.* 2012;4(12):102.

22. Jurtz V, Paul S, Andreatta M, et al. NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J. Immunol.* 2017;199(9):3360–3368.

23. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. *Am. J. Hum. Genet.* 2013;93(4):641–651.

24. Quinn EM, Cormican P, Kenny EM, et al. Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 Genomes data. *PLoS One*. 2013;8(3):e58815.

25. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 2011;39(Database issue):D945-50.

26. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer*. 2016;16(7):413–430.

27. Alsafadi S, Houy A, Battistella A, et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.* 2016;7:10615.

28. Obeng EA, Chappell RJ, Seiler M, et al. Physiologic Expression of Sf3b1K700E Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell*. 2016;30(3):404–417.

29. Darman RB, Seiler M, Agrawal AA, et al. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep.* 2015;13(5):1033–45.

30. Bueno R, Stawiski EW, Goldstein LD, et al. Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. *Nat. Genet.* 2016;48(4):407–416.

31. Hundal J, Carreno BM, Petti AA, et al. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* 2016;8(1):11.

32. Ott PA, Hu Z, Keskin DB, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*. 2017;547(7662):217–221.

33. Sahin U, Derhovanessian E, Miller M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*. 2017;547(7662):222–226.

34. Rodenko B, Toebes M, Hadrup SR, et al. Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* 2006;1(3):1120–1132.

35. Frenkel-Morgenstern M, Lacroix V, Ezkurdia I, et al. Chimeras taking shape: Potential functions of proteins encoded by chimeric RNA transcripts. *Genome Res.* 2012;22(7):1231–1242.

36. Wang X-S, Prensner JR, Chen G, et al. An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. *Nat. Biotechnol.* 2009;27(11):1005–11.

37. Milosevic JD, Puda A, Malcovati L, et al. Clinical significance of genetic aberrations in secondary acute myeloid leukemia. *Am. J. Hematol.* 2012;87(11):1010–6.

38. Wang L, Brooks ANN, Fan J, et al. Transcriptomic Characterization of SF3B1 Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia. *Cancer Cell*. 2016;30(5):750–763.

39. Kahles A, Lehmann K-V, Toussaint NC, et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell*. 2018;

40. Balachandran VP, Łuksza M, Zhao JN, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature*. 2017;551(7681):512.

41. Linette GP, Stadtmauer EA, Maus M V., et al. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood*. 2013;122(6):863–871.

42. Liu S, Tsai W-H, Ding Y, et al. Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res.* 2016;44(5):e47.

43. Abate F, Zairis S, Ficarra E, et al. Pegasus: a comprehensive annotation and prediction tool for detection of driver gene fusions in cancer. *BMC Syst. Biol.* 2014;8(1):97.

Tables

Patients	Count		
Healthy controls	15		
Patient cohort (without replicates)	113		

Demographics			
Sex			
Male	50		
Female	63		
Age, mean +/- SD	66.0 +/- 12.1		

WHO classification (2016)			
ET	32 (30%)		
PV	17 (16.4%)		
PMF	55 (52.9%)		
sAML	9		

Genotype			
ET	17 JAK2-V617F, 11 CALR (type 1 & 2)		
PV	16 JAK2-V617F, 1 JAK2 exon 12		
PMF	26 JAK2-V617F, 29 CALR (type 1 & 2)		
Triple-negative	4		
Post-MPN sAML	8 JAK2-V617F, 1MPL-W515A		

Blood parameters at diagnosis	ET	PMF	PV	
Hemoglobin level	14.0 +/- 1.3	12.6 +/- 1.8	NA	g/dL
White blood cell count	9.4 +/- 2.8	9.9 +/- 5.3	NA	g/dL
Platelet count	853.6 +/- 294.1	762.7 +/- 444.6	NA	$*10^9$/L

Outcome			
Overall survival (months)	185 +/- 77	158 +/- 87	145 +/- 86

Table Legends

Table 1. Description of the patient cohort and clinical parameters. PV patients were all *JAK2* positive (16 *JAK2*-V617F, 1 *JAK2* exon 12), 26 PMF patients were *JAK2*-V617F positive and 29 were *CALR* positive (17 type 1 or type 1-like, 12 type 2 or type 2-like), among ET patients 17 carried *JAK2* mutations, 11 *CALR* mutations (4 type 1, 7 type 2 or type 2-like) and 4 patients were triple-negative cases. All post-MPN sAML patients were *JAK2*-V617F positive except for 1 patient with a *MPL* (W515A) mutation (**Figure 1A, Supplemental Table 1**).

Figure Legends

Figure 1. Large aberrations and small mutations (SNVs and Indels) shape the MPN mutational landscape. (A) Distribution of diagnosis within MPN cohort (outer ring). Occurrence of the MPN driver mutation status within each diagnosis (inner ring). Number of patients for each diagnosis (center). Healthy controls are colored in grey. **(B)** Transcriptome sequencing was performed on granulocyte RNA from 113 MPN and 15 healthy controls. Independent workflows for fusion calling, variant calling (SNVs and Indels) and differential splicing analysis (in *SF3B1* mutated patients) were established. Mutations were called, filtered and validated and those leading to protein changes were translated to amino acid sequences. Neoantigens were predicted on a personalized level, taking each patient's MHCI haplotype into consideration. **(C)** A total of 123 MPN patient samples are depicted and sorted by diagnosis MPN driver mutation status (*JAK2*, *CALR* or *MPL*-positive) and non-driver mutation frequency. Patient sample replicates (n=10) sequenced across or with the same batch (A-E) are indicated in capital letters (A-J). (Chromosomal aberrations panel) Uniparental Disomies (UPDs), deletions and gains were called with Affymetrix SNP 6.0 arrays (**Supplemental Table 5**). (Fusion panel) Fusions were private among patients and colored by rearrangement type. Two fusions were reported for patient P106A#A. For fusion calling we combined the results of 3 fusion detection tools (*defuse*, *soapfuse*, and *tophatfusion*) in order to overcome algorithm-specific biases, a practice frequently suggested in fusion benchmarking studies[42,43]. (MPN driver mutation panel) MPN driver mutation status (*JAK2*, *CALR*, and *MPL*) were determined as described in Supplemental Methods. (SNVs & Indels panel) Genes are grouped by occurrence in pathways and by mutation frequency. Mutations are colored by mutation type and a gradient for high and low SIFT score was applied (only for SNVs). Small rectangles with black frame enhance visibility for mutations with low SIFT score). Only those genes are shown with at least 2 mutations across the cohort (with exception of genes involved in the splicing machinery *SRSF2*, *SF3A1*). Gene names colored in blue are part of the TruSight Myeloid Targeted Sequencing panel. For validation variants from 77 RNA/DNA pairs were compared (**Supplemental Figure 6A-6D**). Out of 113 variants, 91 (81%) were concordant between RNA and DNA, 6 (5%) variants were only called on RNA, and 16 (14%) only on DNA level (**Supplemental Table 10**). Out of the 16 variants called on DNA only, 4 were filtered out due to RNA-specific filters (**Supplemental Methods - RNA-specific filter a-e**) the remaining 12 were not called due to low gene expression and low variant allele frequency.

Figure 2. Recurrent gene mutations, variant allele frequency (VAF) of mutations and gene expression in MPN cohort. Left panel: 35/87 myeloid related genes (**Figure 1C, Supplemental Table 7**) with the highest mutation frequency within cohort are depicted. Mutations are sorted as described in **Figure 1C** and colored by diagnosis. Middle panel: VAF for each mutation colored by mutation type. Right panel: Gene expression values were reported for each patient as fragments per kilobase per million reads (FPKM) (**Supplementary Table 12**). Expression values are plotted as *log2*(FPKM). Patients with a mutation in the respective gene are shown as vertical bars. Numbers indicate the minimum and maximum expression value.

Figure 3. *SF3B1*-mutated PMF patients show distinct aberrant 3' splicing. (A) *SF3B1* mutations were present in 9 PMF patients both, in *JAK2* and *CALR*-positive patients. Other splicing factor mutations (*U2AF1, SRSF2, and SF3A1*) were mutually exclusive to *SF3B1*. **(B)** Gene mutation map of *SF3B1*. Non-synonymous mutations in *SF3B1* were located in oncogenic hotspot sites (K700E and K666N/R/T), except for R594L. Mutations in *SF3B1* were overlaid with mutation data from the COSMIC database[25] (y-axis counts at the bottom). **(C)** Comparison between the frequency of novel 3' and novel 5'splice sites. **(D)** Histogram showing the frequency of the *log2* distance in nucleotides of the aberrant splice site from its canonical site. Zero indicates the start of the exon. **(E)** Hierarchical clustering of *log2*-normalized z-scores for 271 significantly upregulated 3' splice sites in *SF3B1*-positive patients located 10-30 base pairs away from the canonical splice site (**Supplemental Table 14**). **(F)** Mutation-specific splicing patterns were observed for amino acid change K700 and K666.

Figure 4. Percent-splice-in value for aberrant 3' splicing identifies 20 genes with putative alterations on the protein level. (A) Percent-spliced-in (PSI) values for alternative 3' splice sites were calculated by dividing the (a) number of reads spanning the alternative 3' splice site by the (b) number of reads spanning the alternative 3' splice site plus the number of reads spanning the canonical splice site. **(B)** Hierarchical clustering of PSI values of novel 3' splice site versus canonical 3' splice sites of the top 20 genes ranked by ascending PSI values. Only junction sites introducing novel peptide sequences were considered. **(C)** PSI-specific splicing patterns were observed for amino acid change K700 and K666. **(D)** Fragment length analysis was used to validate RNA-

based PSI measurements. **(E)** Validation of PSI values for 4 patients and 3 healthy controls with fragment length analysis. Experimental PSI values were calculated using relative peak heights of canonical and alternate 3' splice sites. No alternative 3' splicing was present in patient and healthy controls for *UBL7*. **(F)** Differential splice junction expression (column 4-5), impact of the splicing aberration on the expression of the affected gene column 7-8). Differential splice junction and gene expression identified 850 and 828 significantly regulated junctions and genes, respectively **(Supplemental Table 13, 17)**. Normalized junction and gene read counts were extracted for the previously identified top 20 genes with highest PSI. Column 6 & 9: *Log2*(Fold change) mutant versus wildtype for junction and gene expression. (Adjusted *P-value*) *P-value*<0.05, ***P-value* <0.0005. Column 10: Predicted amino acid sequences introduced by the splicing defects. Abbr.: chr. – chromosome, Dst. – distal (distal acceptor distance – number or nucleotides between aberrant and canonical splice site).

Figure 5. MPN driver mutations (*CALR & MPL*) and *SF3B1*-mediated splicing aberrations are a rich source of predicted neoantigens. (A) Number of predicted neoantigens (non-synonymous SNVs, Indels and fusions) for each patient, based on individual HLA haplotypes. For the prediction the 4 most common alleles for *HLA-A*, *HLA-B* and *HLA-C* within the MPN cohort were considered (**Supplemental Table 19**). Thresholds for weak (%Rank <2 & >0.5) and strong binders (%Rank <0.5) were selected based on the NetMHCpan authors' recommendation. **(B)** Percentile rank score (%Rank) values for *CALR*, *MPL,* and **(C)** *SF3B1* peptides predicted with NetMHCpan are plotted. *HLA* alleles, marked in red, were used for validation (**Figure 5F**). HLA MPN cohort allele frequencies are indicated in figure legend of **Figure 5C (D)** Number of predicted neoantigens (all mutation classes combined) for each patient with no *CALR* or *SF3B1* mutation, either of the two, or mutations in both genes, separated into weak and strong binders. *CALR* mutated patients had a mean of 8.0 weak binders and 2.3 strong binders, whereas *SF3B1* mutated patients had an average of 38.2 weak binders and 16.2 strong binders without co-occurring with *CALR* mutations. For *JAK*2-V617F mutated patients, we found an average number of 0.9 and 0.1 predicted neoantigens for weak and strong binders, respectively (**Supplemental Table 23**). **(E)** *CALR* mutant tail "consensus sequence" as reported in Klampfl et al.[1]. The dots indicate the starting positions of peptides colored by peptide length. **(F)** Selected peptides (**Supplemental Table 24**) were synthesized and tested for binding to 4 MHCI molecules using an ELISA-based peptide

exchange assay. The threshold for predicted binding was set as explained for **Figure 5A**. For experimental binding validation, the threshold was set at 2 standard deviations (2SD) from the mean of absorbance values of 3 positive control peptides (colored in red). Peptides above this threshold were considered strong binders. Of the 20 predicted strong binders, 15 gave binding values above the threshold (75.0%). Of the 15 predicted weak binders, 8 were above the threshold (53.3%). Color coding: green – *CALR*, yellow rectangle – *MPL,* and magenta – *SF3B1* derived peptides. Abbreviations are listed in **Supplemental Table 24**.

Figure 1

Figure 2

Figure 3

Figure 4

A

5'ss 3'ss

A   B

$$PSI = \frac{a}{a+b}$$

B

Batch: A B C

Diagnosis: PMF, Control

Genotype: CALR–type1, CALR–type2, JAK2–V617F, wildtype

SF3B1 status: *SF3B1* mut, *SF3B1* wt

Batch
Diagnosis
Genotype
SF3B1 status

mut PSI
1
0.8
0.6
0.4
0.2
0

UBL7
SIRPD
TTI1
EFEMP2
MAP3K7
OXA1L
FLT3LG
VWA7
CBY1
ZFYVE27
PRPF38A
PPM1M
HINT2
RWDD4
C7orf43
CC2D1A
TBRG1
ZNF397
TMCC2
ACOT11

PSI - Percent splice in
0   25   50   75   100

C

K700E
K700E
K700E
K700E
K666N
K666N
K666T
K666R

D

MAP3K7
P069A K700E
175   210

Relative fluorescent units (RFUs)
28000
21000
14000
7000
0

3'ss
alt 3'ss

180   200 nts

TBRG1
P069A K700E
210   245   280

12000
9000
6000
3000
0

3'ss
alt 3'ss

235   248 nts

E

P009A_K666N	P040B_K666N	P048A_K666T	P069A_K700E	C1	C2	C3	
0	0	0	0	0	0	0	*UBL7*
18	35	33	37	0	0	0	*SIRPD*
46	49	78	56	0	11	0	*TTI1*
74	100	83	75	0	23	0	*EFEMP2*
50	50	89	67	19	22	15	*MAP3K7*
49	58	56	30	3	3	2	*OXA1L*
34	34	32	26	20	19	17	*FLT3LG*
19	41	55	38	0	0	0	*VWA7*
34	19	29	36	0	0	0	*CBY1*
7	0	13	35	0	0	0	*ZFYVE27*
14	7	20	39	0	0	0	*PRPF38A*
28	19	47	46	0	2	0	*PPM1M*
34	38	33	23	4	2	6	*HINT2*
40	32	50	45	25	0	5	*RWDD4*
45	35	25	27	2	5	0	*C7orf43*
40	37	35	38	9	0	0	*CC2D1A*
24	23	50	25	0	0	0	*TBRG1*
50	27	41	35	6	26	0	*ZNF397*
50	46	45	0		0	8	*TMCC2*
19	0	0	55			0	*ACOT11*

mut PSI (%)
0   100
NA

F

Gene name	Acceptor chr. coordinates	Dst. acceptor distance	Junction expression			Wildtype gene expression			Novel peptide sequence
			SF3B1 mut	*SF3B1* wt	log2(*SF3B1* mut/wt)	*SF3B1* mut	*SF3B1* wt	log2(*SF3B1* mut/wt)	
UBL7	15:74740929	29	4.95	2.17	2.28	10.33	10.14	-0.08	SHTPTPGTQGPHQCPLVSSQGRPSPMISSAKPYSMPFRPLGSPAFRASGSPSCSSYVTWASRTMS*
* *SIRPD*	20:1517975	15	2.33	0.66	3.51	7.31	7.33	-0.54	GRGTQVFVTVLPLSAEQNPRPPKNR
TTI1	20:36631196	17	4.44	1.56	2.84	8.53	8.83	-0.38 *	NVSDFDNEEVVLFFQRNSQSLPKWMRMTPVQMWSHHCHCRSK*
EFEMP2	11:65635912	19	6.49	2.02	3.21	7.81	7.91	-0.62	QLLATRLCQAPLPWHRH*
MAP3K7	6:91269954	20	7.89	3.50	2.25	10.22	10.31	-0.08	YAEGGSLYNVCAFLSQCCMVLNHCHIILLPTQ*
OXA1L	14:23238985	14	11.39	5.38	2.12	11.53	10.73	0.86 ***	LPWWGAIAAFCIFQVQSLPAA*
FLT3LG	19:49978930	17	5.29	4.29	1.23	4.31	4.45	-0.09	VLAPAWSPTHPLPCSLSPPAAAAELGTQWDPGLLLPTQPHLLRLRCQNP*
VWA7	6:31741234	19	2.83	0.88	3.21	6.78	6.96	-0.44	FGTHPPKPPVPFPNPGKCSHGGHFD
CBY1	22:39066874	14	2.08	0.53	3.97	5.64	5.87	-0.13	SASLSNLHSHLAHSWIDQPGRWSWAWNTDPRLSESTAESHLMEKELDELRISRKRK*
ZFYVE27	10:99504468	17	4.14	0.78	5.33	10.96	10.95	-0.25	NVLFLTLNEVFLILQVHGTQ*
* *PRPF38A*	1:52880412	21	5.88	1.71	3.43	10.43	10.69	-0.15	RSRSPKRRSLTSLFYSPSPRRERHR
PPM1M	3:52283671	14	9.96	2.10	4.76	12.39	12.26	-0.25	PGNQEDPHRMVFTGSQSWPRC*
HINT2	9:35813154	11	2.70	0.97	2.78	5.49	4.71	0.70	EGLGDGYRLAQFQ*
RWDD4	4:184577128	13	3.67	0.55	6.64	8.15	8.38	-0.19	MSANEDQEFFISDGTRSITLYL*
C7orf43	7:99752805	17	10.48	3.81	2.75	12.44	11.93	-0.25	QQPSRSHLMSLCLPPGRAV*
* *CC2D1A*	19:14034130	15	4.66	0.62	7.50	9.93	9.52	0.17	NGLPVDITKPRPPQVPPAPVNKD
TBRG1	11:124496355	13	5.23	0.39	13.38	10.53	10.91	-0.38	KGKENNKLEDLRFSSEENMQEKENGGRCSQAGSAHCPGSLRTACVPHRTRGSNSI*
* *ZNF397*	18:32834174	21	3.34	0.68	4.89	10.22	10.33	-0.18	PQEPSFRGITFPFCFLEELSRPPKASS
TMCC2	1:205240923	17	6.81	2.47	2.76	9.00	6.43	2.41 *	SYERARDIQPLCPCRRPWSPA*
ACOT11	1:55100337	23	1.56	0.42	3.71	6.96	6.15	1.27	LSTLKAIPVRLCLRFSRPKSTADLDTFKTPETLS*

* in frame

Normalized read count
0   12

Normalized read count
4   13

69

Figure 5

Supplemental Methods

Mutational Landscape of the Transcriptome Offers Putative Targets for Immunotherapy of Myeloproliferative Neoplasms

Unique patient and sample identifiers

Unique patient and sample identifier start with capital letter 'P', control samples with a 'C', followed by a three digit number unique to each patient. After the digits a letter code [A-Z] identify individuals blood samplings (A..first sampling, B…second sampling etc). The '#' separates the patient information (left of '#') from the batch information (right of '#'). Batches are labelled from A to E. (Example: P026A#B – Patient P, number 026, first blood sampling A, sequenced in batch B)

Sample replicates

Samples from 7 PMF, 2 PV, and 1 ET patient/s of the MPN cohort were sequenced in duplicates across different batches or within the same batch (see library preparation) as technical replicates. For patient P060 two serial samples (A and B) were drawn 6 years apart.

Data integration and batch description

Samples were labelled with different batch names A, B, C, D, and E and differed in RNA isolation procedure, library preparation, sequencing protocols, and the person handling the sample. In this study we integrated patient transcriptome sequencing data from these 5 different sequencing experiments (**batch A-E**) (**Supplemental Table 2**).

Microarray analysis

DNA of 122/124 patient samples were processed and hybridized to Genome-Wide Human SNP 6.0 arrays (Affymetrix) according to manufacturer's instruction. Arrays were analyzed as described in Klampfl. et al., 2011 (1).

Fusion transcript detection with RNA sequencing data

Raw FASTQ reads were processed with defuse (release 0.6.2) (2), tophatfusion (release 2.0.14) (3) and soapfuse (release 1.26) (4) using the following parameters [**defuse**: *defuse.pl –c $config_file -1 $R1_fastq -2 $R2_fastq –o $output_file*; **tophatfusion**: *tophat2 –o $output_directory --fusion-search --keep-fasta-order --bowtie1 --no-coverage-search –r $inner_mate*

_distance –mate-std-dev 80 –max-intron-length 100000 –fusion-min-dis 100000 –fusion-anchor-length 13 –fusion-ignore-chromosome chrM $genome_reference $R1_fastq $R2_fastq_; **soapfuse**: _SOAPfuse-RNA.pl –c $config_file –fd $sample_directory –l $execution_directory –o $output_directory_]. For defuse and soapfuse ENSEMBL human genome GRCh37 (version 69) (http://www.ensembl.org/info/ website/archives/) and for tophatfusion UCSC human genome (version hg19) (http://hgdownload. soe.ucsc.edu/goldenPath/hg19/) was used.

The fusion candidates obtained by defuse, tophatfusion and soapfuse were combined and further filtered to reduce the amount of false positives calls using an in-house python script released on github (https://github.com/sp00nman/rnaseqlib). The workflow and filtering included (1) for defuse a minimum score of >0.8 and for tophatfusion >0 was necessary (soapfuse does not provide scores). (2) [**healthy**] removal of fusion candidates found in at least one sample out of 25 healthy granulocyte RNA samples (3) [**read evidence**] fusion predictions required a minimum amount of 2 reads covering the junction of a fusion breakpoint and 1 read pair flanking the fusion breakpoint (4) [**read-through, dshort distance**] fusions between two neighboring genes or two genes in close proximity (less than 1000 base pairs apart) were removed only if an inversion or eversion event could be excluded (5) [**no protein**] predictions had to involve at least one protein coding fusion partner (6)[**false positive**] fusions were excluded where the fusion partner genes were totally or partially overlapping (7) [**pseudogenes**] as well as fusions between paralog or pseudogenes (8) [**mappability**] mappability tracks from ENCODE (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncode Mappability/) were used to extract mappability scores of 50 base pairs around the fusion breakpoint for each gene. Fusion genes were one of the partner gene had a mappability score less than 0.1 were excluded (9) [**blacklisted region**] fusion genes were the breakpoint falls within regions of DAC (www.encodeproject.org/annotations/ENCSR636HFF/) were removed. (10) [**HB gene, HLA gene**] Finally, fusion candidates involving highly polymorphic genes of the human leukocyte antigen (HLA, http://hla.alleles.org/genes/) and homologous hemoglobin genes (HBB, http://www.ensembl.org/) were excluded from analysis. A detailed description of the filters applied are summarized on _github_ (https://github.com/sp00nman/rnaseqlib/blob/master/ doc/fusion_workflow.md).

Fusion genes that passed the above filters were further filtered manually (sequencing reads were inspected with an IGV genome browser) . Filtered fusions were flagged with the following tags _(mappability, missed_readthrough, non-uniform splitread distribution, missed_healthy, max 5 fusions/sample, DGV)_. (11) _Mappability_ refers to failed primer

design due to lack of uniquely mapping primers around the fusion breakpoint. (12) *Missed_read-through* are fusions that through manual inspection with an IGV genome browser have a high likelihood of being fusions mediated through splicing rather than chromosomal aberrations (short distance; median distance of 50kb, same strand, no gene in between). (13) *Non-uniform splitread distribution* are fusions where the breakpoint of the split reads cluster at the beginning or end of the read. (14) *Missed healthy* are fusions that due to false ID mapping were not filtered out in the automatized filtering workflow. (15) *Max 5 fusions/sample* filter was applied only to tophatfusion and only two samples (P003B#A, P091A#A) with an unusual high number of fusions detected. (16) *DGV* flags refer to fusions where the fusion partner genes were located on the same chromosome and the breakpoints are within 100kb away from CNV breakpoints reported by the Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/docs/DGV.GoldStandard. July2015.hg19.gff3).

The abundance of fusion transcripts was estimated with SRPKM (seed RPKM), which is calculated by extracting the split and spanning reads and normalizing by library size and read length. *(#seed reads * 10^3 * 10^6)/(read_length * #mappable reads)* as proposed by *FusionMap*(5).

Assessing sensitivity of fusion calling workflow

The sensitivity of the fusion detection workflow was assessed with a total of 20 fusions provided by MLL München Laboratory (GmbH, München) and sequenced in-house. The raw reads were processed as previously described.

Cytogenetics and FISH

Chromosome preparations and banding analysis were performed as previously described according to standard methods (6). If required, cases with an aberrant karyotype in chromosome banding analysis were additionally investigated by 24-color FISH to completely resolve the karyotype (24XCyte Human Multicolor FISH Probe Kit, Metasystems; Altlussheim, Germany). For classification of abnormalities and karyotypes, the ISCN guidelines (2016) were used (7). Interphase FISH (fluorescence in situ hybridization) was performed to determine chromosomal rearrangements of the respective genes/loci. The following probes were applied: *BCR-ABL1, CBFB-MYH11, ETV6, IGH, IGH-BCL2, IGH-CCND1, IGH-FGFR3, IGH-MYC, MECOM, PDGFRB, PML-RARA, RUNX1-RUNX1T1* (all purchased from Metasystems; Altlussheim,

Germany), *CRLF2* (Cytocell, Cambridge, UK), *FIP1L1-CHIC2-PDGFRA* (Kreatech; Amsterdam, Netherlands), *KMT2A* (Abbott; Wiesbaden, Germany).

Sanger sequencing validation of fusion genes

The High-capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA) was used for RT-PCR according to the manufacturer's protocol. Primer sequences flanking the fusion breakpoints are listed in **Supplemental Table 4**. The PCR products were Sanger sequenced using the BigDye Terminator version 3.1 cycle-sequencing kit (Life Technologies) and analyzed on a 3130xl Genomic Analyzer (Applied Biosystems, Foster City, CA). Sequences were analyzed with CLC Genomics Workbench (Qiagen, release 4.9) to define the exact fusion junction site and the frame of the novel fusion transcript. Validated fusion candidates were screened in up to 10 healthy individuals and excluded if these occurred in any of those 10 controls.

Mapping genomic breakpoints of the fusion junctions

The genomic breakpoint for validated fusions were identified by PCR using primer pair combinations derived from the introns flanking the putative fusion junction. The products of positive PCR reactions were Sanger sequenced to determine the fusion breakpoint at the genomic level.

Variant calling on RNA sequencing data

Raw reads were aligned with Spliced Transcripts Alignment to a Reference (STAR) (release 2.4.2a) (8) using the following options *[-outFilterIntronMotifs RemoveNoncanonical*; *-outSAMtypeBAM SortedbyCoordinate*] followed by STAR-2-pass method as described here (8). Human reference sequence b37 was used to align the short reads. Duplicate reads were removed with Picard (*MarkDuplica*tes; version 1.118; https://github.com/broadinstitute/picard). GATK tool *SplitNCigarReads* was used to hard-clip sequences overhanging into the intronic regions. Indel realignment and base quality recalibration was performed as recommended for variant calling on DNA level. GATK *HaplotypeCaller* was used for calling variant with RNA specific options [-*dontUseSoftClippedBases*; *-recoverDanglingHeads*; *-stand_call_conf 20.0*]. Hard cutoff filters were set with GATK *VariantFiltration* tool to reduce false positive calls. These filters included (a) removal of occurrence of at least 3 SNVs that are within a window of 35 bases (SnpCluster), (b) removal of variants with a fisher strand value of more than 30 and (c) a quality by depth of less than 2, based on GATK best practice workflow for RNA variant calling (9). Additionally, (d) variants that fall within or are 1bp away from a

homopolymer stretch of at least 5 nucleotides were removed as well as (e) SNVs that were located within 5 bases away from an Indel. Remaining variants were annotated with ANNOVAR (release 2015-06-17) (10).

Variant filtering criteria

Variant calling using RNA sequencing bears a number of challenges. (a) Using RNA as the underlying source for variant calling means that any gene that is not expressed cannot be called (b) GATK's Haplotype Caller uses algorithms to calculate variant likelihoods that have difficulties dealing with extreme frequencies, therefore capturing variants in small sub-clones (<0.1 variant allele frequency) without prior knowledge of the mutation is not feasible. (c) The absence of matching germline samples renders difficult a precise distinction between somatic and germline mutations. We therefore established a set of stringent filtering criteria with the aim to enrich for somatic variants: (a) We excluded variants annotated as 'common' in the Single Nucleotide Polymorphism Database dbSNP (version 142), (b) variants that were annotated with a minor allele frequency (MAF) of more than 0.01 by the 1000 Genomes Project (version 2015-02) in one of the following populations (EUR, AMR, EAS, SAS, AFR), (c) variants with a MAF>0.01 found in NHLBI Exome Sequencing Project (ESP) (version esp5400_all & esp6500siv2_all) (d) we further filtered for lowly expressed variants and required a variant to have an alternate read count of at least 3 unless a COSMIC (version 84) entry was found for that variant, (e) variants were excluded which appeared in 1/15 healthy controls (in-house data) but a minimum variant allele frequency of 0.4 was required, (f) variants that were annotated as synonymous SNVs were disregarded and finally, (g) we extracted all exonic variants of 87 genes (**Supplemental Table 7**), of these 54 were part of the TruSight Myeloid Sequencing Panel (Illumina, San Diego, CA), the remaining were genes with prior implications in myeloid-proliferative disorders as well as candidate genes from overlapping regions of common aberrations (UPD, deletion, gain) (1,11).

Gene mutation maps for frequently mutated genes

Gene mutation maps were drawn using *lollipops* (release 2015) (**Figure 3B, Supplemental Figure 7**). Protein domains were extracted from the Pfam database. Mutation data from different cancers were downloaded from the COSMIC database (12).

Targeted re-sequencing of patients' genomic DNA

Samples were processed using the TruSight Myeloid Sequencing Panel which allowed for quantitative evaluation of mutations in 54 genes implicated in myeloid diseases (**Supplemental Table 7**). In addition, we generated and sequenced a single amplicon covering two *SF3B1* hotspot mutation sites (K666 and K700). PCR was performed using Herculase Fusion Polymerase (Agilent, Santa Clara, CA) applying standard conditions as recommended by the manufacturer at 55° annealing temperature, using primer pools (**Supplemental Methods Table 1**).

Supplemental Methods Table 1. Primer pool sequence.

Primer name	Primer sequence
SF3B1_K666_K700_F0	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-GAGTTGCTGCTTCAGCCAAG
SF3B1_K666_K700_F1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNGAGTTGCTGCTTCAGCCAAG
SF3B1_K666_K700_F2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNGAGTTGCTGCTTCAGCCAAG
SF3B1_K666_K700_F3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNGAGTTGCTGCTTCAGCCAAG
SF3B1_K666_K700_R0	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG(staggers)GCAAAAGCAAGAAGTCCTGG
SF3B1_K666_K700_R1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNGCAAAAGCAAGAAGTCCTGG
SF3B1_K666_K700_R2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNGCAAAAGCAAGAAGTCCTGG
SF3B1_K666_K700_R3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNGCAAAAGCAAGAAGTCCTGG

Primers included staggers required for amplicon diversity and resulting amplicons comprised primer binding sites required for sample indexing using the Nextera XT Index Kit v2 (Sets A-D, 384 indexes; Illumina San Diego, CA). Primers were pooled equimolar and used at standard concentrations.

DNA isolated from granulocytes or peripheral blood mononuclear cells (PBMCs) was processed according to the manufacturer's instructions to generate indexed amplicon-based libraries. Equimolar amounts of libraries were pooled into multiplexes (96-plexes for TruSight, 384-plexes for the single-amplicon approach) which were then sequenced 150bp paired-end on an Illumina HiSeq3000 or MiSeq instruments for the TruSight Myeloid Panel and the single-amplicon approach, respectively. Read alignment and variant calling was performed using the BaseSpace software (Illumina, San Diego, CA). Variants called in transcribed regions or at splice sites were selected and further filtered for common variation as described in section *Variant calling on RNA sequencing data*.

Other filters were adjusted for TruSight targeted sequencing and included insufficient sequencing read depth (<200) and low variant allele frequency (VAF <0.05).

Reporting on RNA & DNA concordance

To estimate the quality of detecting somatic SNVs from transcriptome sequencing data we compared genomic targeted sequencing data from a myeloid gene panel of 54 genes (**Supplemental Table 6**) for 77 patients (**Supplemental Table 9**) of our RNA sequenced cohort. For RNA, we only considered variants that were located within exons part of the myeloid gene panel. For DNA, we compared variants with a variant allele frequency >= 0.16 which is the lowest observed RNA variant allele frequency after filtering (**Supplemental Table 10**).

Estimating expression values on a gene level and differential expression analysis

Sequencing reads were trimmed for quality and adapter sequence content with Trimmomatic in single-end (ILLUMINACLIP: TruSeq3-SE.fa:2:30:10:1:true, SLIDINGWINDOW:4:15, MINLEN:20) or paired-end (ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10:1:true, SLIDINGWINDOW:4:15, MINLEN:20) mode. The resulting reads were aligned with the STAR aligner (https://www.ncbi.nlm.nih.gov/ pubmed/23104886) to the UCSC hg38 reference genome resembling the Genome Reference Consortium GRCh38 assembly, with the basic Ensembl transcript annotation set from version 87 (December, 2016). Reads overlapping transcript features were counted with the *summarizeOverlaps* function of Bioconductor library *GenomicAlignments*, taking into account that the Illumina TruSeq stranded mRNA protocol leads to sequencing of the second strand so that all reads need inverting before counting. The Bioconductor package *DESeq2* was then used to model the data set and call differentially expressed genes. The design matrix included modelling both, *batch* and *gender* as additive effects: *~ batch + sf3b1 + sex*

Definition of novel 3' splice junctions

Novel splice junctions were defined as (a) not present in GENCODE (version 14), (b) having at least 20 reads covering that particular splice junction over all samples, (c) sharing a known 5' or 3' splice site annotated in GENCODE and finally, (d) followed the canonical splice acceptor/donor motifs: GU/AG, CU/AC, GC/AG, CU/GC, AU/AC, GU/AU. . Differential splice site usage was tested with the *testForDEU* function of the *DEXSeq* Bioconductor package (version 1.14.2, R version 3.2.3 2015-12-10).

Fragment length analysis using capillary electrophoresis

Aberrant 3' splice sites were amplified by RT-PCR using 5'FAM labelled forward primers in combination with reverse primers (**Supplemental Table 16**). PCR products were diluted 1:20 in HPLC-grade water (Sigma-Aldrich) and sent to Microsynth (Microsynth AG Switzerland) for Fragment Length Analysis. Peak calling was performed with GeneMapper 4.0 software (Applied Biosystems). PSI scores were calculated using the fractions of peak height for canonical and alternate 3' splice sites as previously explained.

Filtering of NetMHCpan results

We filtered NetMHCpan results based on the following criteria: (a) We selected peptide:MHC-binding based on percentile rank score (%Rank) predictions (%Rank scores reflect the relative affinity of a peptide compared to a pool of random peptide sequences), which were <2 for weak and %Rank<0.5 for strong binders (author's recommendation), (b) non-synonymous SNVs with a wildtype sequence %Rank <2 were excluded, (c) an FPKM of >0 for variants and a SRPKM >0 for fusions was necessary for further evaluation, (d) comparison of predicted peptide to human protein sequences, (e) for splice factor mutations we selected all aberrant 3' splice sites 10-30 nucleotides away from its canonical splice site (previous publication have identified AG nucleotides used as aberrant splice sites to be positioned about 13-17 base pairs downstream of the branch point, but more than 10 base pairs upstream of the canonical splice in order to avoid competition for splice sites (13,14)), (f) selected splice sites with a minimum median PSI of 50% in mutant *SF3B1* PMF patients and less than 20% median PSI in *SF3B1* wildtype patients.

HLA-typing & validation

HLA alleles from the patient cohort were extracted in 4-digit resolution using *seq2HLA*(15) (release 2.2) using RNA-seq data as input (*P value* threshold of 0.01) (**Supplemental Table 18**) . To validate the RNA based *HLA-A*, *HLA-B*, *HLA-C* allele calls additional genotyping by standard DNA based assays of 8 *SF3B1* mutated patients was performed at the Department for Blood Group Serology and Transfusion Medicine of the Medical University of Vienna (1090 Vienna, Austria). We could validate 49/50 alleles on a 2-digit resolution and 47/50 on a 4-digit resolution.

Generate FASTA files with peptide sequences for each mutation class

For SNVs we built amino acid *FASTA* sequences using 9 amino acids flanking each side of the mutated variant. Similar *FASTA* sequences were generated for the wildtype sequence allele. For Indels, we only considered mutations that cause shifts in the reading frame. Novel reading frames were translated until the first stop codon was encountered. The entire stretch of novel amino acid sequences was written to *FASTA* files. For fusions with the potential formation of novel fusion proteins, we extracted 9 amino acids flanking the fusion breakpoint. Finally, aberrant 3' splicing events which caused frameshifts, were treated as previously described. For *in frame* splicing, we built *FASTA* sequences using 9 amino acids flanking the novel peptide stretch. We used *biomaRt* (R Bioconductor, version 2.34.0) as a resource for annotation, sequence extraction and translation as well as custom R scripts for converting the variants into *FASTA* sequences.

References

1.   Klampfl T, Harutyunyan A, Berg T, Gisslinger B, Schalling M, Bagienski K, et al. Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. Blood. 2011;118:167–76.

2.   McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. Defuse: An algorithm for gene fusion discovery in tumor rna-seq data. PLoS Comput Biol. 2011;7.

3.   Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. BioMed Central Ltd; 2011;12:R72.

4.   Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol. 2013;14:R12.

5.   Ge H, Liu K, Juan T, Fang F, Newman M, Hoeck W. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. Bioinformatics. 2011;27:1922–8.

6.   Schoch C, Schnittger S, Bursch S, Gerstner D, Hochhaus A, Berger U, et al. Comparison of chromosome banding analysis, interphase- and hypermetaphase-FISH, qualitative and quantitative PCR for diagnosis and for follow-up in chronic myeloid leukemia: a study on 350 cases. Leukemia. 2002;16:53–9.

7.   McGowan-Jordan J, Simons A, Schmid M. An International System for Human Cytogenetic Nomenclature (2016). Karger. 2016.

8.   Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, Rätsch G, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods. 2013;10:1185–91.

9.   Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome

Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;11:11.10.1-11.10.33.

10. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.

11. Lundberg P, Karow A, Nienhold R, Looser R, Hao-Shen H, Nissen I, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood. American Society of Hematology; 2014;123:2220–8.

12. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011;39:D945-50.

13. DeBoever C, Ghia EM, Shepard PJ, Rassenti L, Barrett CL, Jepsen K, et al. Transcriptome Sequencing Reveals Potential Mechanism of Cryptic 3' Splice Site Selection in SF3B1-mutated Cancers. Wang E, editor. PLOS Comput Biol. Public Library of Science; 2015;11:e1004105.

14. Darman RB, Seiler M, Agrawal AA, Lim KH, Peng S, Aird D, et al. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. Cell Rep. Elsevier; 2015;13:1033–45.

15. Boegel S, Löwer M, Schäfer M, Bukur T, de Graaf J, Boisguérin V, et al. HLA typing from RNA-Seq sequence reads. Genome Med. 2012;4:102.

16. Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC, et al. Somatic *CALR* Mutations in Myeloproliferative Neoplasms with Nonmutated *JAK2*. N Engl J Med. Massachusetts Medical Society ; 2013;369:2391–405.

Supplemental Figure Legends

Supplemental Figure 1. Bioinformatics analysis workflow. We established data processing and filtering workflows for fusion calling, variant (SNVs & Indels) calling, differential splicing analysis, gene expression analysis, HLA typing and neoantigen discovery. Validation steps were included for each of the workflows (excluding gene expression analysis).

Supplemental Figure 2. Assessing the sensitivity of the fusion calling workflow. **(A)** The sensitivity of the fusion detection workflow was assessed by applying RNA-sequencing on 20 different hematological tumors with fusions validated with standard karyotype analysis or fluorescence in situ hybridization (FISH). Three representative examples of FISH validated fusions are shown. **(B)** In 15/20 patient samples a candidate fusion was identified. Fusions where the driver gene partner contributes only to the regulatory sequence of the fusion gene (for example IGH-rearrangements [IGH-X]) were missed in 6/20 samples. Exceptions are *IGH-CRLF2* rearrangements that were identified in 3 samples. Typical acute myeloid leukemia (AML) related fusions like *PML-RARA*, *MLL-X*, *RUNX1-RUNX1T1*, *CBFB-MYH11* were detected by all fusion detection tools (*soapfuse*, *defuse*, *tophatfusion*). Fusion expression is reported as seed reads per kilobase per million reads SRPKM (min-max [1.6-57.1] SRPKM).

Supplemental Figure 3. Fusion filtering for each patient of MPN cohort. MPN patients and healthy controls were processed through the same fusion calling workflow as samples described in **Supplemental Figure 2**. Fusions detected with **(A)** *defuse,* **(B)** *soapfuse,* **(C)** *tophatfusion* and a score greater than 0.8 (*defuse*), none (soapfuse), and 0 (*tophatfusion*) are plotted. For each sample the total amount of fusions, filtered fusions, and the filter tag each fusion received is reported. **(D)** A total of 138, 204, and 94 fusions for *defuse*, *soapfuse,* and *tophatfusion*, respectively, passed all applied filters and were subject to **(E)** manual inspection. **(E)** 37, 20, and 46 fusions were considered for validation of which **(F)** 20, 3, and 22 fusions were experimentally verified by RT-PCR followed by Sanger sequencing. These verified fusions were screened for appearance in cDNA in up to 10 healthy individuals. **(G)** Finally, a total of 5, and 3, and 11 fusions for *defuse*, *soapfuse,* and *tophatfusion* were considered high confidence fusions. **(H)** The small overlap of fusions detected by different detection tools is a known limitation among RNA-seq specific fusion detection software 4–6 and was also observed in our dataset with

only 1 fusion shared among all tools and 5 fusions among 2 tools. The union of all fusion detection tools resulted in a total of 13 unique fusions.

Supplemental Figure 4. Sanger validation of fusions. The exact breakpoint region was verified on cDNA level with Sanger sequencing. For fusion **(A)** *BCR-ABL1*, **(B)** *XPO5-RUNX2*, and **(C)** *KDM4B-CYHR1* the cDNA breakpoint is located at exon boundaries suggesting the location of genomic breakpoint to be intronic. **(D)** *FRY-MYCBP2*, **(E)** *CUX1-intragenic*, **(F)** *AHCTF1-SMYD3* (H) *INO80D-GPR1* are examples of fusions with a 5' cDNA breakpoint location within an exon boundary and a 3' cDNA breakpoint either intronic, genomic or located within an exon. Splice acceptor and donor sites (marked in red rectangles) are present. In fusion **(G)** *NOX5-PNISR* both cDNA breakpoints are located within introns. Reciprocal transcript was confirmed with Sanger sequencing. Fusions **(J)** *B2M-SELL*, **(K)** *CSF3R-ZFP36*, **(L)** *SAT1-MNDA*, **(M)** *S100A8-PLEK*, **(N)** *S100A9-CTSD* have repetitive regions of 5-7 nucleotides overlapping the cDNA breakpoint located within exons, UTR regions, or intronic regions. Splice acceptor and donor sites are absent.

Supplemental Figure 5. Analysis of co-occurring fusions and variants with MPN driver mutations. (A) Circos plot summarizing aberrations detected with SNP 6.0 arrays (outer circle) overlaid with location of cDNA fusion breakpoints. Aberrations are colored by aberration type. Fusions are colored by diagnosis (inner circle). **(B)** Number of fusions for each rearrangement type. **(C)** Fusion expression is reported as seed reads per kilobase per million reads SRPKM (min-max [1.6-57.1] SRPKM), which is calculated by extracting the split and spanning reads and normalized by library size and read length. **(D)** Wildtype gene expression of upstream and downstream gene of fusion in log2(FPKM). Rearrangement types are indicated in symbols.

Supplemental Figure 6. Sensitivity of RNA variant calling workflow. Cumulative distribution of variant allele frequency from 77 patients for variants (SNV & Indels) called on **(A)** DNA, **(B)** RNA level and colored by mutation type. **(C)** Correlation of *JAK2*-V617F burden measured on RNA level versus allele specific PCR. **(D)** Percentage concordance for DNA and RNA for SNVs as a function of RNA coverage with varying thresholds for variant allele frequency.

Supplemental Figure 7. Gene mutation maps for frequently mutated genes in MPN. Position of mutations along the gene are shown. Protein domains & regions were

extracted from Pfam and are marked in grey. Detected mutations were overlaid with frequency mutation data from COSMIC database.

Supplemental Figure 8. Hierarchical clustering of differentially expressed splice junctions. Hierarchical clustering of 271 differentially expressed splice junctions of 87 MPN patients.

Supplemental Figure 9. Comparison of 3' splice site usage among 7 diseases. (A) Heatmap of Percent splice in (PSI) values ranked by PSI of 271 differentially expressed novel 3' splice sites (**Supplemental Table 14**). **(B)** Comparison of PSI values for top the 20 genes with highest PSI value. **(C)** Overlap of differentially spliced 3' junctions between combinations of sets for 7 diseases. Plot was drawn with R package *UpsetR* (version 1.3.3). **(D)** Description of publications and diseases used for comparison. Numbers indicate the number of patients and or samples used for differential junction expression for each disease.

Supplemental Figure 10. Number of predicted neoantigens for each patient in MPN cohort. Neoantigens are colored by MHCI molecule they are predicted to bind to (upper panel), peptide length (middle panel), and mutation type (lower panel).

Supplemental Table Legends

Supplemental Table 1. Patient and clinical data.

Supplemental Table 2. Sample processing and sequencing information.

Supplemental Table 3. FISH validated fusions and reported fusions using in-house fusion detection workflow. (related to **Supplemental Figure 2**)

Supplemental Table 4. Fusions detected with defuse, soapfuse & tophatfusion. Remaining fusions after applying filter 1-10 (Supplemental methods). Fusions were further manually filtered (filter 11-16) (column "filter.manual"). Fusions that passed the manual filter were validated with Sanger sequencing using primers listed in column "forward.primer" & "reverse.primer" (related to **Figure 2A, Supplemental Figure 5B, 5C**).

Supplemental Table 5. Chromosomal aberrations detected with SNP 6.0 arrays for each patient. (related to **Supplemental Figure 5B**)

Supplemental Table 6. Wildtype gene expression values (FPKM) for upstream and downstream gene of fusion for each patient sample. (related to **Supplemental Figure 5D**)

Supplemental Table 7. List of 87 genes used for RNA variant calling. Column "trusight.selection" indicates whether the gene is part of the TruSight Myeloid targeted re-sequencing panel (Illumina).

Supplemental Table 8. List of variants (SNVs & Indels) called on RNA level. (related to **Supplemental Figure 5A**)

Supplemental Table 9. List of patients and diagnosis used for comparison of RNA and DNA variants. (related to **Supplemental Figure 6**)

Supplemental Table 10. List of variants with concordant calls from RNA and DNA variant calling. Column "validation" indicates whether the variant was a concordant call, only called on RNA or only called on DNA level. (related to **Supplemental Figure 6**)

Supplemental Table 11. Mutation frequency in total number or as percentage for each disease entity. Distinctions were made whether the variant was reported as "somatic" in COSMIC database. Gene frequencies were compared to reported gene frequencies in Nangalia et al., 2013 (Nangalia et al., 2013) (related to **Figure 2B**).

Supplemental Table 12. Wildtype gene expression values (FPKM) for each sample in the cohort for 35 frequently mutated genes in MPN. (related to **Figure 2B**)

Supplemental Table 13. 850 differentially spliced junctions between *SF3B1* mutated and wildtype patients. Only junctions with *P-value*<0.1 and log2(FC)>0 were included.

Supplemental Table 14. 271 differentially spliced junctions between *SF3B1*-mutated and wildtype patients with aberrant 3' splicing. Only aberrant 3'splicing 10-30 nucleotides from canonical splice site were included (related to **Figure 3C, 3D, 3E**)

Supplemental Table 15. Comparison of differentially spliced junctions across 7 diseases merged from 4 publication. Publications are indicated in the column "publication" (related to **Supplemental Figure 9**).

Supplemental Table 16. 43 novel 3' splice junctions with a median PSI > 20% for mutated *SF3B1* patients. (related to **Figure 4B, 4F**)

Supplemental Table 17. 828 differentially expressed genes of *SF3B1* mutated versus non-mutated PMF patients & controls. (related to **Figure 4F**)

Supplemental Table 18. HLA genotypes extracted from RNA with *seq2HLA* for each patient.

Supplemental Table 19. HLA allele count and frequency for MPN cohort. Table is sorted by *HLA-A*, *HLA-B* and *HLA-C* allele and then by frequency. (related to **Figure 5**)

Supplemental Table 20. NetMHCpan peptide:MHCI binding predictions for variants. Predictions for variants called on RNA level for the four most common alleles within the RNA cohort. Results are shown for each sample (related to **Figure 5A**).

Supplemental Table 21. NetMHCpan peptide:MHCI binding predictions for *MPL*. (related to **Figure 5B**)

Supplemental Table 22. NetMHCpan peptide:MHCI binding predictions for aberrantly spliced genes (in *SF3B1* mutated patients). (related to **Figure 5C**)

Supplemental Table 23. Number of predicted neoantigens for each patient. Weak and strong binders are indicated in separate columns (related to **Figure 5D**).

Supplemental Table 24. Comparison of experimental (peptide exchange assay) versus predicted binding (NetMHCpan). (related **to Figure 5F**)

Supplemental Figure 1

Process/Analyse

Filter

Validation

Granulocyte RNA-seq

FASTQ

Fusion calling

Process reads with:
defuse (release 0.6.2)
tophatfusion (release 2.0.14)
soapfuse (release 1.26)

Removal of fusion candidates found in at least 1 out of 15 healthy controls (granulocyte RNA)

Minimum number of supporting reads
split reads >= 2
span reads >= 1

Tool specific filtering thresholds or scores for
defuse (p.val) >0.8
tophat (score) >0
soapfuse (not provided)

Fusions formed by readthrough events (neighboring genes)

One fusion partner is required to be protein coding

Totally or partially overlapping genes, paralog & pseudogenes

Fusion partner genes require a mappability score of >0.1

Fusions with breakpoints falling within regions of DAC

Fusions involving highly poly-morphic genes (*HLA*,*HBB*)

Manual inspection with IGV browser

Sanger sequencing of fusion candidates

Validated fusions were screened in up to 10 healthy individuals and removed if present

Variant calling & differential splicing

Map to reference b37 + decoy with *STAR* (release 2.4.2a) + star2pass alignment

SJ.out.tab

Remove duplicate reads *Picard* (release 1.118), *Mark Duplicates*

Hard-clip sequences over-hanging into intronic regions *GATK* (release 3.4/46), Split'N'Trim

Indel realignment - *GATK*

Base recalibration - *GATK*

Variant calling with *GATK*, *Haplotype Caller* HC in RNAseq mode

Removing SNVs within a SNV Cluster (>=3 SNVs within a window of 35 base pairs

Remove variants with fisher strand (FS) value >30

Remove variants with a quality by depth (QD) <2

Flag and remove variants within 5bp from Indel (SNV near Indel)

Flag and remove variants within or near (~1bp) homopolymer runs (>= 5 nucleotides)

Annotation with ANNOVAR (release 2015-06-17)

Keep variants with an alternate read count of at least 3 (unless a COSMIC (version 84) exists

Remove variant if it appears in at least 1 samples out of 15 healthy samples (1 sample required VAF>0.4)

Select putative deleterious variants (non_synonymous, frameshift_deletions, stopgain, stoploss)

Exclude variants with MAF >0.01 according to databases (dbSNP common, 1000 Genome Project, Exome Variant Server)

For 77 samples targeted sequencing of matching DNA samples for 54 genes

Enrich for somatic variants

Call novel splice junctions: According to DeBoever et al. & custom scripts

Novel junctions defined as: Not present in GENCODE (version 14)

At least 20 reads covering the particular splice junction over all samples

A known 5' or 3' splice site needs a GENCODE (version 14) annotation

Novel junctions defined as...

Test for differential juction expression with *DEXSeq* (version 1.14.2), *testForDEU*

Signficant 3' splice sites with a mutant Percent splice in (PSI)>0.2 and a wildtype PSI<0.2

Abberrant splice site in protein coding region

Fragment length analysis with capillary electrophoresis

Expression analysis

Reads were trimmed for quality and adapter sequence content with *Trimmomatic*

Map to reference UCSC hg38 with *STAR* (release 2.4.2a) + star2pass alignment

Read counting with Bioconductor library *GenomicsAlignments* function *summarizeOverlaps*

HLA-typing

HLA alleles calculated using *seq2HLA* (release 2.2)

9 *SF3B1* mutated patients were genotyped at the Departement for Blood Group Serology and Transfusion Medicine

Neoantigen discovery

Generate FASTA file with peptide sequences (8-10 mers) (custom script)

Predict Peptide:MHC binding affinity with NetMHCpan (version 4.0)

Keep all peptide binding predictions with a %Rank <2

For nonsynonymous SNVs wildtype peptide should have a %Rank >2

Gene expression FPKM>0

Peptide:MHC binding Peptide:MHC UV-exchange assay

Supplemental Figure 2

A

B

Supplemental Figure S3

Supplemental Figure 4 (1/4)

Protein domains
- cDNA breakpoint
- exon (E) boundaries
- splice acceptor/donor site

Sanger sequencing track - cDNA
- coding sequence
- intronic/genomic sequence

A BCR chr 22 (+) | ABL1 chr 9 (+)

Translocation

Protein domains

BCR domains
① Bcr-Abl_Oligo
② RhoGEF domain
③ RhoGAP domain
④ C2 domain

ABL1 domains
① SH3_1
② SH2 domain
③ Protein tyrosine kinase
④ F-actin binding

Sanger sequencing track - cDNA
BCR [22:23,632,600] - ABL1 [9:133,729,451]

B XPO5 chr 6 (-) | RUNX2 chr 6 (+)

Inversion

Protein domains

XPO5 domains
① Xpo1

RUNX2 domains
① Runt domain
② Runxl

Sanger sequencing track - cDNA
XPO5 [6:43,527,915] - RUNX2 [6:45,459,678]

C KDM4B chr 19 (+) | CYHR1 chr 8 (-)

Translocation

Protein domains

KDM4B domains
① JmjN
② JmjC domain, hydroxylase
③ PHD_2
④ PHD-zinc-finger-like domain

CYHR1 domains
① Zinc finger, TRAF-type

Sanger sequencing track - cDNA
KDM4B [19:5,047,680] - CYHR1 [8:145,678,843]

89

Protein domains
- cDNA breakpoint
- exon (E) boundaries
- splice acceptor/donor site

Sanger sequencing track - cDNA
- coding sequence
- intronic/genomic sequence

D FRY chr 13 (+) | MYCBP2 chr 13 (-)

Deletion

13

FRY ▸ ◂ MYCBP2

5' fusion
3' fusion

chr 13q deletion

FRY(+)   MYCBP2(-)

chr 13 - FRY region

FRY   NRCAM NMP3L PDS5B

chr 13 - MYCBP2 region

MYCBP2

Protein domains

E 2, 90 aa

FRY ① ② ③ ④ ⑤ ⑥ 3011aa

MYCBP2 ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ 4679 aa

I2, 29aa

FRY domains
① Cell morphogenesis N-terminal
② Cell morphogenesis central region
③ MOR2-PAG1_mid
④ MOR2-PAG1_mid
⑤ Cell morphogenesis central region
⑥ Cell morphogenesis C-terminal

MYCBP2 domains
① RCC1_2
② RCC1
③ PHR domain
④ PHR domain
⑤ Filamin/ABP280 repeat
⑥ Bacterial SH3 domain
⑦ Anaphase-promoting complex, subunit 10 (APC10)
⑧ zf-RING_2

Sanger sequencing track - cDNA
FRY [13:32,653,170] - *MYCBP2* [13:77,899,452]

```
FRY (E2,E3)       TCATTATGGCAGAGCCCCTGGAAAAGCCATTGACAAAATC
FRY (E2,I2)       TCATTATGGCAGAGCCCCTGgtgagtaatagaaaataacc
MYCBP2 (rev compl I2)  tctttaactcattttgacagtttcctactatgtttttcccc
Sanger track      TCATTATGGCAGAGCCCCTG              tgtcaaagctggtga
                  I  M  A  E  P  L  F  P  T  M  F  S     C  Q  S  W  *
```

PCTARSCGKKKMKPTPNYH

E CUX1 chr 7 (+) | intragenic chr 7 (-)

Inversion

13

CUX1

5' fusion
3' fusion

Protein domains

E 1, 10aa

CUX1 ① ② ③ ④ 1512aa

CUX1 domains
①
② CUT domain
③
④ Homeobox

Sanger sequencing track - cDNA
CUX1 [7:101,459,373] - *intragenic* [7:148,174,607]

```
TTGATTTACAGCAGCTGCAG AGAGAACTCGATGCCACCGC CUX1 (E1, E2)
TTGATTTACAGCAGCTGCAG gtcaggctcctccgcgctcg CUX1 (E1, I1)
atgtgagatttccatgacag gttgtatctgtttgctcggg rev genomic
TTGATTTACAGCAGCTGCAG gttgtatctgtttgctcggg Sanger track
 D  L  Q  Q  L  Q  V  S  V  C  S  G  CHNKGPQTGGFKQQKFTFSQYWRLEV*
```
+1

F AHCTF1 (ENST00000478568.1) chr 1 (-) | SMYD3 chr 1 (-)

Inversion

1

5' fusion
3' fusion

SMYD3 AHCTF1

Protein domains

E 1

AHCTF1 ① 67aa

SMYD3 ① ② 428aa

I 5

AHCTF1 domains
① beta-propeller of ELYS nucleoporin

SMYD3 domains
① zf-MYND
② SET domain

Sanger sequencing track - cDNA
AHCTF1 [1:247,094,880] - *SMYD3* [1:246,167,120]

5'UTR AHCTF1   start codon
```
CGCTCACCCGCCCCGTTCAG TGGAAGT ATGCGAGACTTAA  AHCTF1 (E1, E2)
CGCTCACCCGCCCCGTTCAG gtggttcgcgtcctgcgccg  AHCTF1 (E1, I1)
ttgcctttctttctccacag ttttcccatgtgattctctt  SMYD3 (rev compl I5)
CGCTCACCCGCCCCGTTCAG ttttcccatgtgattctctt  Sanger track
```
alternative start codon?

Protein domains
- cDNA breakpoint
- exon (E) boundaries
- splice acceptor/donor site

Sanger sequencing track - cDNA
- coding sequence
- intronic/genomic sequence

Protein domains

G INO80D chr 2 (-) | GPR1 chr 2 (-)

Tandem duplication

2

- 5' fusion
- 3' fusion

INO80D   GPR1

INO80D   INO80D-GPR1   GPR1

Protein domains

E 8

INO80D ① ② 1027aa

GPR1 ① 355aa

E 4, cDNA breakpoint
nucleotide position 1333

INO80D domains
① DNA-binding domain
② DNA-binding domain

GPR1 domains
① 7 transmembrane receptor (rhodopsin family)

Sanger sequencing track - cDNA
INO80D [2:206,882,404] - *GPR1* [2:207,041,275]

```
AAGAACATGCCAAAAAAATGGATAATTTCTTGAGAGGAGA  INO80D(E8,E9)
AAGAACATGCCAAAAAAATGgtaagttctgcaacagaggc  INO80D(E8,I8)
ACTTGTGTCTCATCTTCAAGGTGAAGAAGCGAAGCATCCT  GPR1(E4, cDNA bp 1333)
AAGAACATGCCAAAAAAATGGTGAAGAAGCGAAGCATCCT  Sanger track
        E  H  A  K  K  M  V  K  K  R  S  I
```

+1 +1 *inframe*

H NOX5 chr 15 (+) | PNISR chr 6 (-)

Translocation

6

PNISR

NOX5-PNISR

NOX5

PNISR-NOX5

- 5' fusion
- 3' fusion

15

E 1

NOX5 ① ② ③ ④ ⑤ 765aa

PNISR ① 805aa

I4
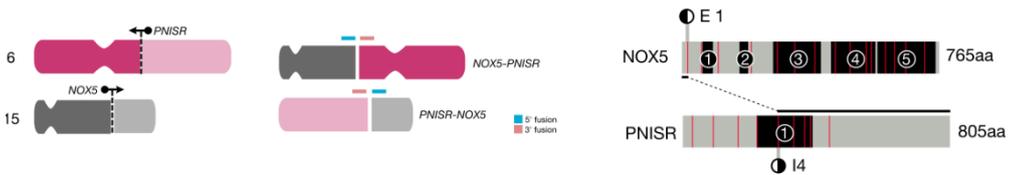
NOX5 domains
① EF-hand_6
② EF hand
③ Ferric reductase like transmembrane component
④ FAD-binding domain
⑤ Ferric reductase NAD binding domain

PNISR domains
① PNISR

Sanger sequencing track - cDNA
NOX5 [15:69,266,540] - *PNISR* [6:99,859,884]

```
tatgtatcctcttgaggttggctttttttttttttttcact  NOX5 (I1)
tatgatgtagatttcttgatgctaattcagtcttccatgg   PNISR (rev I4)
tatgtatcctcttgaggttggctaattcagtcttccatgg   Sanger track
```

PNISR [6:99,859,888] - *NOX5* [15:69,266,355]

```
cagtatgatgtagatttccttagtagctaattcagtcttcc  PNISR (rev I4)
agcttcatgtaaccattaccacaatcaagatgcttaactc   NOX5 (I1)
cagtatgatgtagatttcttacaatcaagatgcttaactc   Sanger track
```

I INO80D chr 2 (-) | GPR1 chr 2 (-)

Sanger sequencing track - genomic DNA
INO80D [2:206,584,716] - *GPR1* [2:206,750,070]

```
ACTTAGAACAGAGATGCTCAGTAAGTAATTGTTGAATA  INO80D(I8)
TTATATATTGTTTGGCTTTTGTTCTGGGAATTCCAGGA  GPR1(E4)
ACTTAGAACAGAGATGCTCTGTTCTGGGAATTCCAGGAA  Sanger track
```
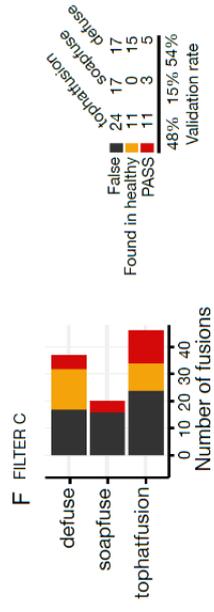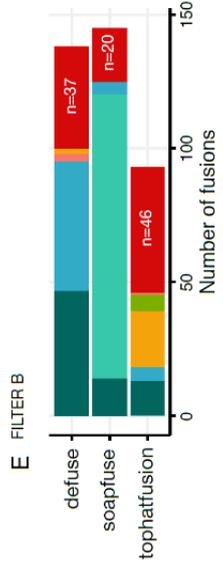
Supplemental Figure 4 (4/4 continued)

Translocations

J B2M chr 15 (+) | SELL chr 1 (-)

K CSF3R chr 1 (-) | ZFP36 chr 19 (+)

L SAT1 chr X (+) | MNDA chr 1 (+)

M S100A8 chr 1 (-) | PLEK chr 2 (+)

N S100A9 chr 1 (+) | CTSD chr 11 (-)

Protein domains   Sanger sequencing track - cDNA

cDNA breakpoint
exon (E) boundaries
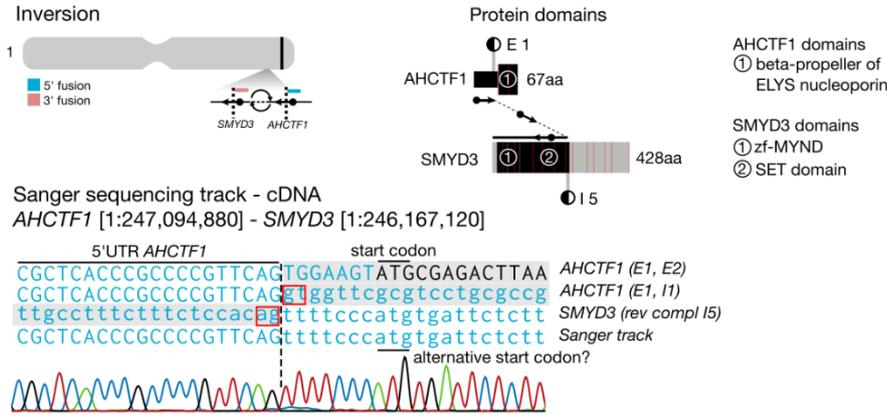splice acceptor/donor site

coding sequence
intronic/genomic sequence

Sanger sequencing track - cDNA

B2M [15:45,010,014] - SELL [1:169,660,280]

CSF3R [1:36,932,399] - ZFP36 [19:39,898,660]

SAT1 [X:23,804,191] - MNDA [1:158,813,166]

S100A8 [1:153,362,596] - PLEK [2:68,619,597]

S100A9 [1:153,333,277] - CTSD [11:17,78,707]

Supplemental Figure 5

A

Diagnosis (inner circle)
- sAML
- ET
- PV
- PMF

Aberrations (outer circle)
- Deletion
- Amplification
- Uniparental Disomy (UPD)

B

Translocation
Inversion
Deletion
Duplication

Number of fusions

C

FRY-MYCBP2
CUX1-7:148,174,607
BCR-ABL1
XPO5-RUNX2
INO80D-GPR1
PNISR-NOX5
S100A8-PLEK
AHCTF1-SMYD3
SAT1-MNDA
CSF3R-ZFP36
S100A9-CTSD
B2M-SELL
KDM4B-CYHR1

in frame
Nonsense-mediated decay (NMD) (predicted)

SRPKM

D

Wildtype gene expression log2(FPKM)

Inversion
Translocation
Duplication
Deletion
5'gene
3'gene

S100A9_CTSD
S100A8_PLEK
B2M_SELL
CSF3R_ZFP36
SAT1_MNDA
PNISR_NOX5
AHCTF1_SMYD3
KDM4B_CYHR1
FRY_MYCBP2
CUX1_chr7:148174607
BCR_ABL1
INO80D_GPR1
XPO5_RUNX2
NOX5_PNISR

Supplemental Figure 6

Supplemental Figure 7

A TET2

K147X
T497S P570A Q769X Q960X Y1255X C1271fs L1678fs M1701I(x3)
       C784S Y867H G986fs C1298fs H1380P Q1680X V1718L(x2) R1878P
                     Q769X   K1422fs   M1907K

Core catalytic region
2002 aa

B DNMT3A

E30A R326C S393fs R729L R736H R790W R882H
                                          R882P
912 aa
PWWP Methyltranferase

C SF3B1

R594L K666T(x1) K700E(x3)
       R(x1)/N(x2)
SF3b1 HEAT domain
1304 aa

D U2AF1

Q157P
Zn Zn
240aa

E CUX1

P604S A742T S1134C K1184fs
CUT CUT CUT homeobox
1505 aa

Number of mutations (COSMIC)

Mutation type
frameshift deletion   nonsynonymous SNV
frameshift insertion   stopgain

95

Supplemental Figure 8

Supplemental Figure 9

Supplemental Figure 10

3.3 Clinical relevance of canonical *SF3B1* mutations in MPN

To further assess the frequency of *SF3B1* mutated patients in an independent cohort, we genotyped additional 482 MPN patients (161 ET, 208 PV, 113 PMF) using a targeted resequencing approach. *SF3B1* hotspot mutations (K666 & K700) were found in 10 PMF and 2 PV patients (12/482 (all MPN) = 2.5%, 10/113 (PMF only) = 8.9%. Overall occurrence for PMF was 18/168=10.7% (both cohorts combined) (**Table 3**).

Table 3: *SF3B1* mutated patients in MPN cohort. Data was pooled from several sequencing experiments indicated in column 'Detection method'. *TruSight* and *SF3B1_targ_seq* are targeted sequencing approaches performed on DNA, *RNA-seq* summarizes all *SF3B1* mutated patients from the RNA sequenced cohort. Column 'VAF-DNA' and 'VAF-RNA' is the mutant allele frequency extracted from DNA or RNA, respectively. Samples P009A, P040B, P055C, P058A, and P069A were sequenced on RNA and DNA level. Abbr.: Progression assoc. – Progression association, adv. PMF – advanced PMF, Ref – Reference allele, Alt – Alternate allele, aa_change – amino acid change, UPN – Unique patient identifier

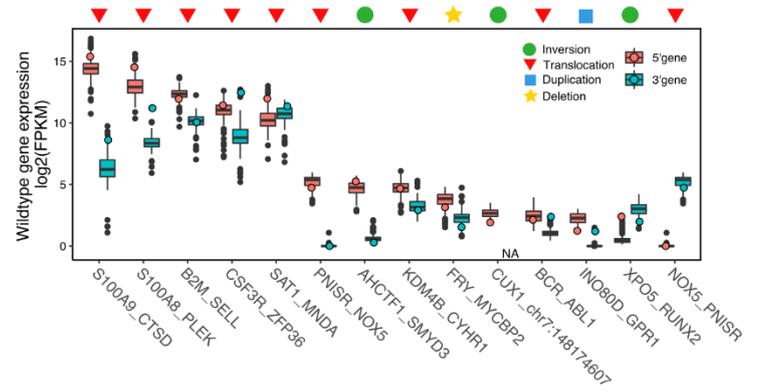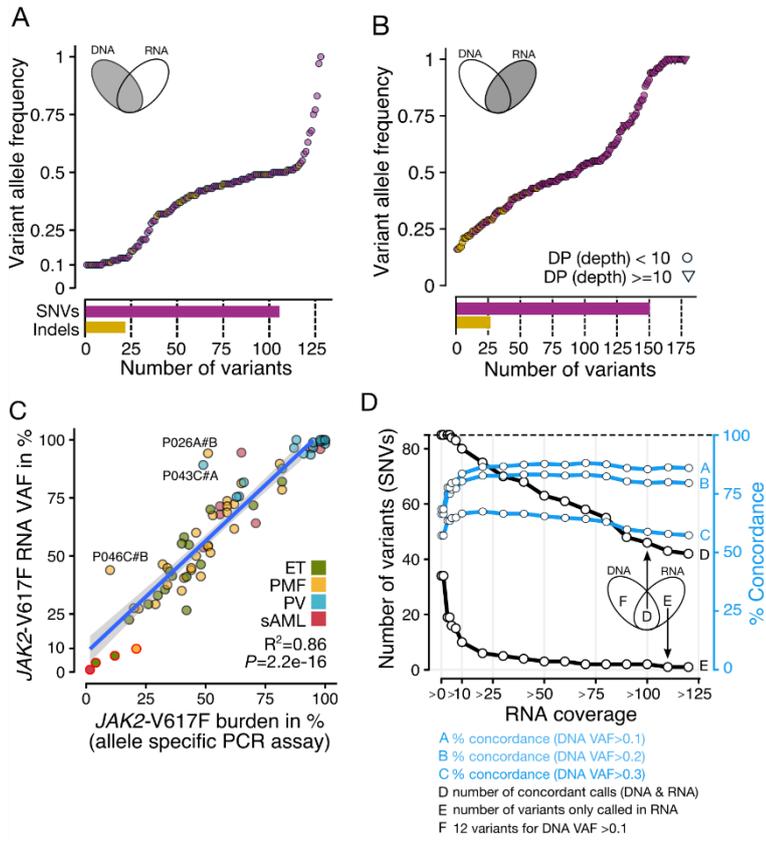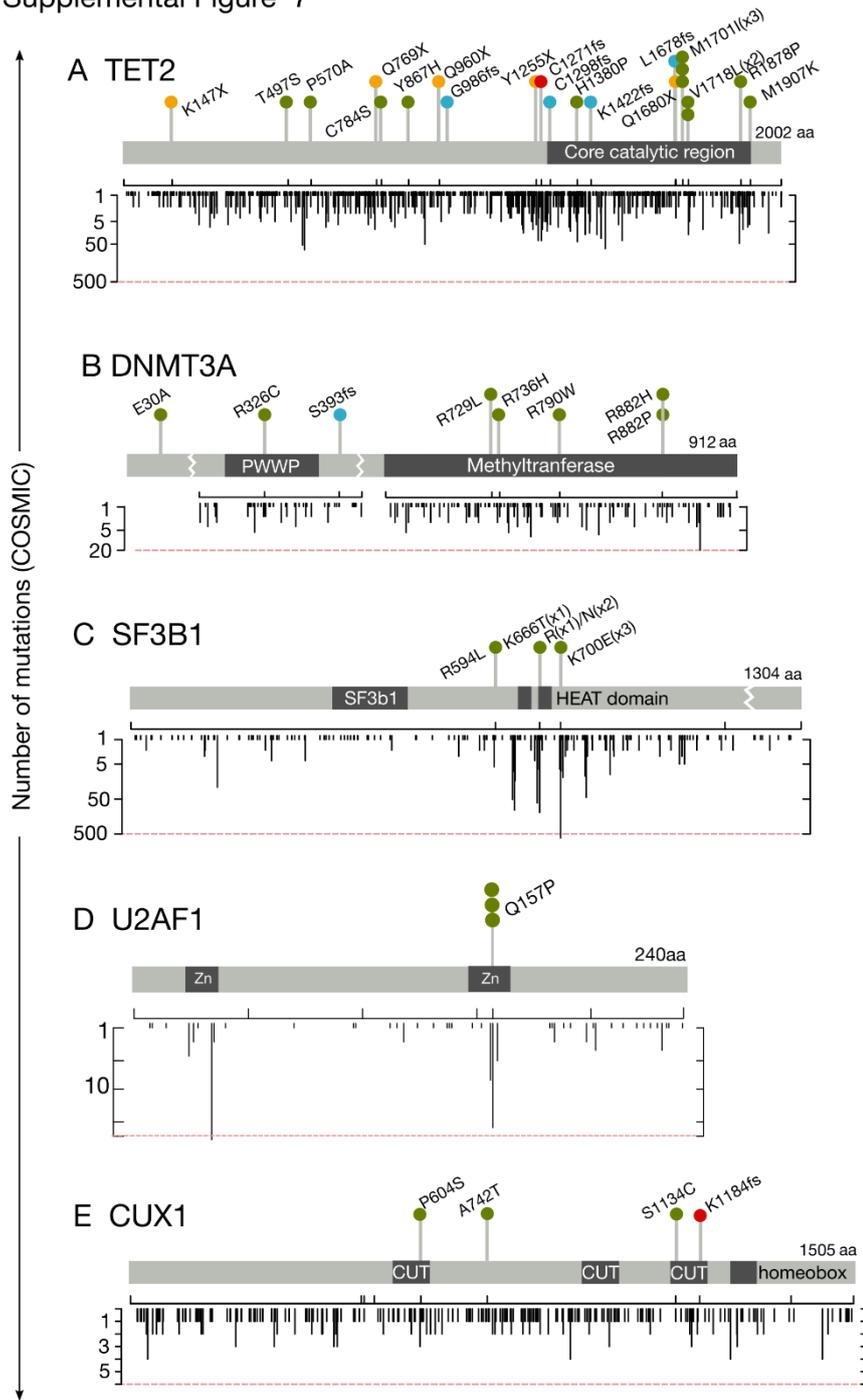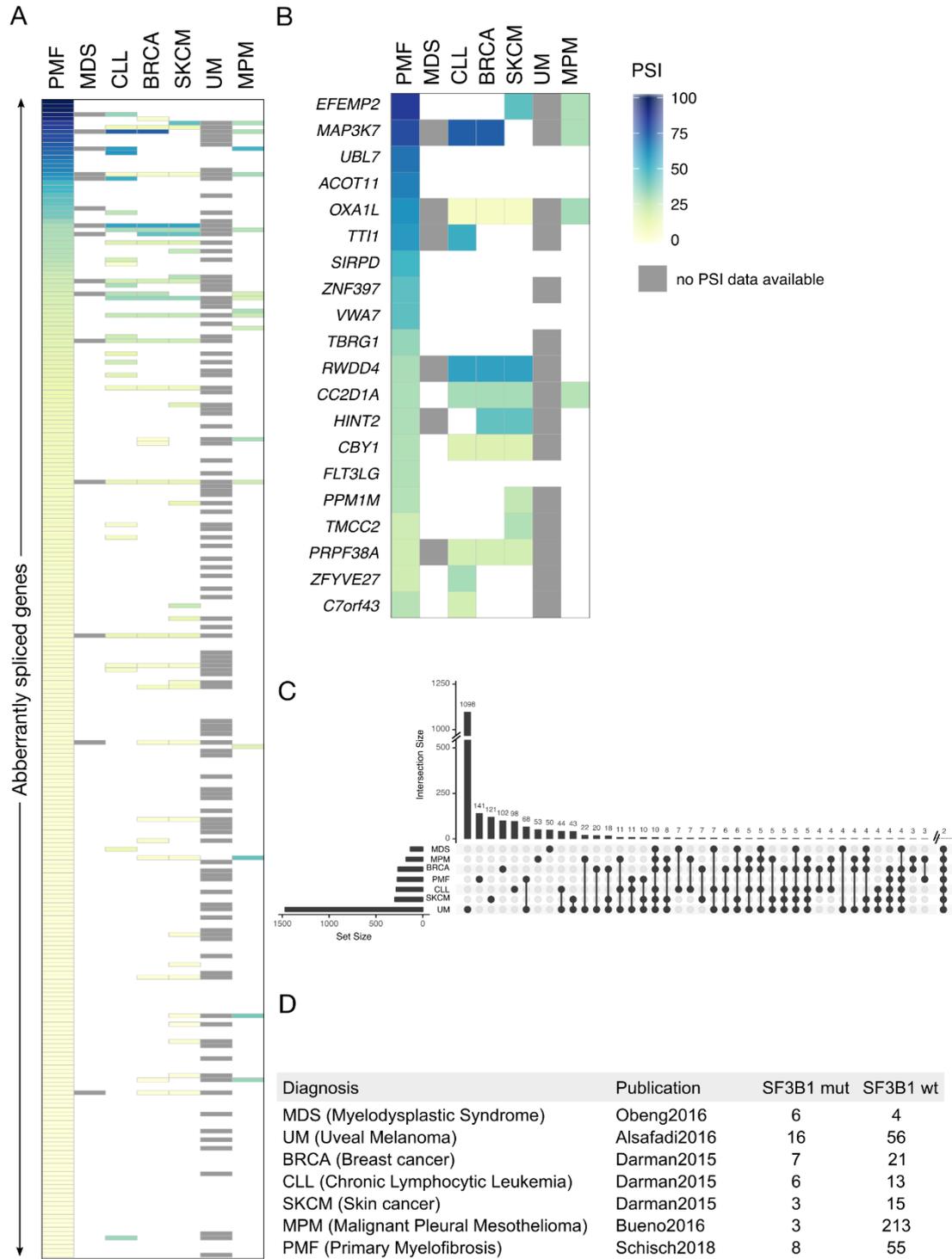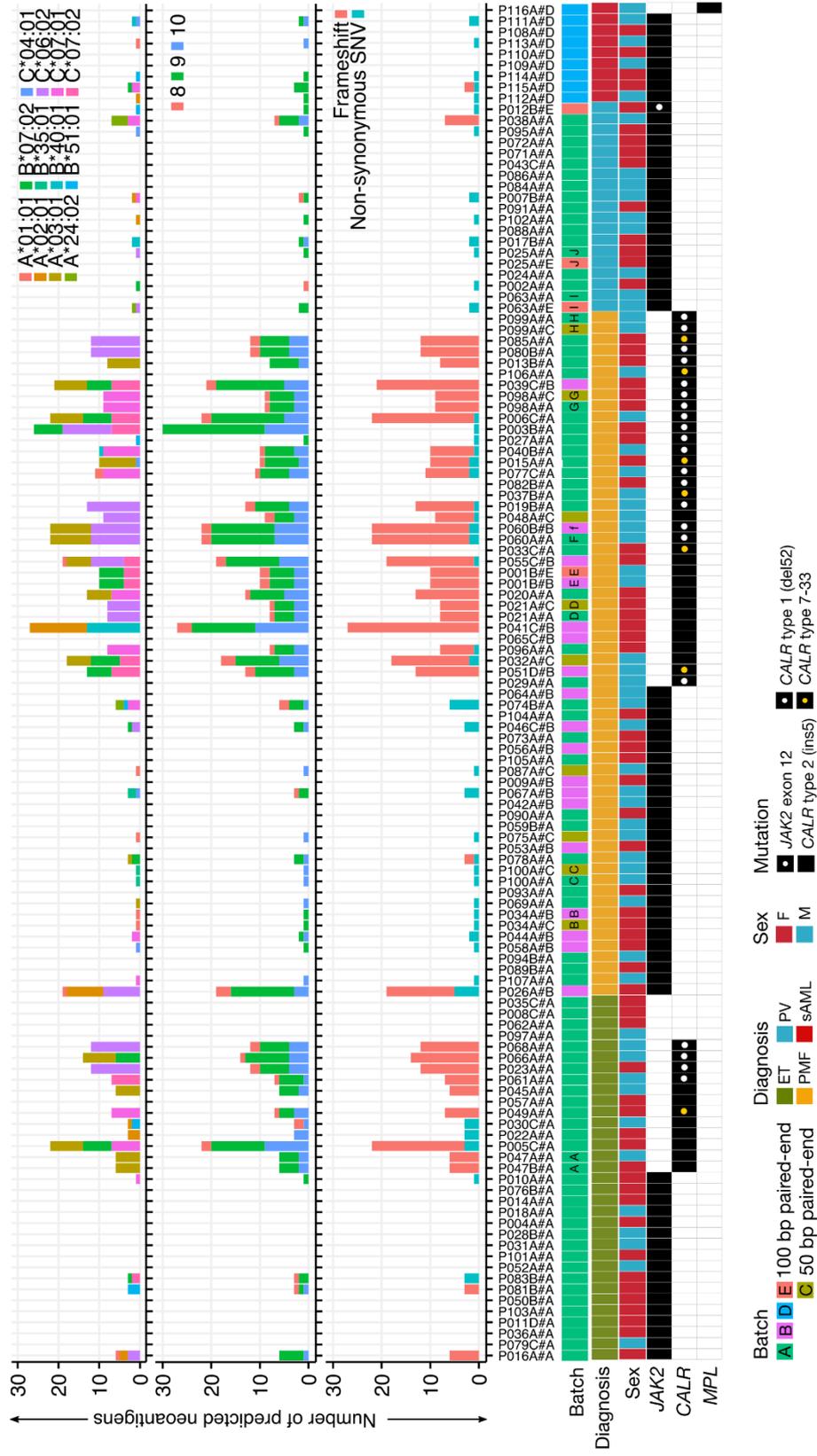| UPN	Diagnosis	aa_change	VAF_DNA	VAF_RNA	Position	Ref	Alt	Detection method	Progression assoc.
001PMF	PMF	K666N	49%		2:198267359	C	G	TruSight	adv. PMF
002PMF	PMF	K666N	47%		2:198267359	C	A	TruSight	adv. PMF
003PMF	PMF	K666N	17%		2:198267359	C	G	TruSight	
004PMF	PMF	K666N	21%		2:198267359	C	G	TruSight	
005PMF	PMF	K666N	37%		2:198267359	C	G	TruSight	
006PMF	PMF	K666N	47%		2:198267359	C	G	TruSight	
007PMF	PMF	K700E	42%		2:198266834	T	C	TruSight	
008PMF	PMF	K700E	52%		2:198266834	T	C	SF3B1_targ_seq	
009PMF	PMF	K700E	3%		2:198266834	T	C	SF3B1_targ_seq	post-PV MF
010PMF	PMF	K700E	31%		2:198266834	T	C	SF3B1_targ_seq	
P009A	PMF	K666N	49%	48%	2:198267359	C	G	RNA-seq	adv. PMF
P034A	PMF	K700E		43%	2:198266834	T	C	RNA-seq	
P040B	PMF	K666N	42%	55%	2:198267359	C	A	RNA-seq	
P048A	PMF	K666T		40%	2:198267360	T	G	RNA-seq	adv. PMF
P055C	PMF	K700E	21%	44%	2:198266834	T	C	RNA-seq	
P058A	PMF	K700E	47%	46%	2:198266834	T	C	RNA-seq	
P064A	PMF	K666R		26%	2:198267360	T	C	RNA-seq	
P069A	PMF	K700E	32%	46%	2:198266834	T	C	RNA-seq	AML
001PV	PV	K666N	49%		2:198267359	C	A	SF3B1_targ_seq	
002PV	PV	K666N	18%		2:198267359	C	A	SF3B1_targ_seq	

Out of 18 PMF patients, 4 had progressed to advanced forms of the disease (accelerated phase), 1 patient had progressed from a previous PV, and 1 patient had transformed to AML. We hypothesized whether *SF3B1* hotspot mutations could be associated with disease progression in MPN. The following analyses were performed to support this observation. First, we wanted to test whether *SF3B1* mutated and post-MPN sAML patients cluster closely and distinctly from wildtype patients. We performed unsupervised principal component analysis of the top 500 most variable splice sites for our entire cohort of 113 chronic and post-MPN sAML patients and observed optimal disease separation in

PC1 and PC3 explaining 26.7% and 7.0% of total variance, respectively (**Figure 17 & 18**).
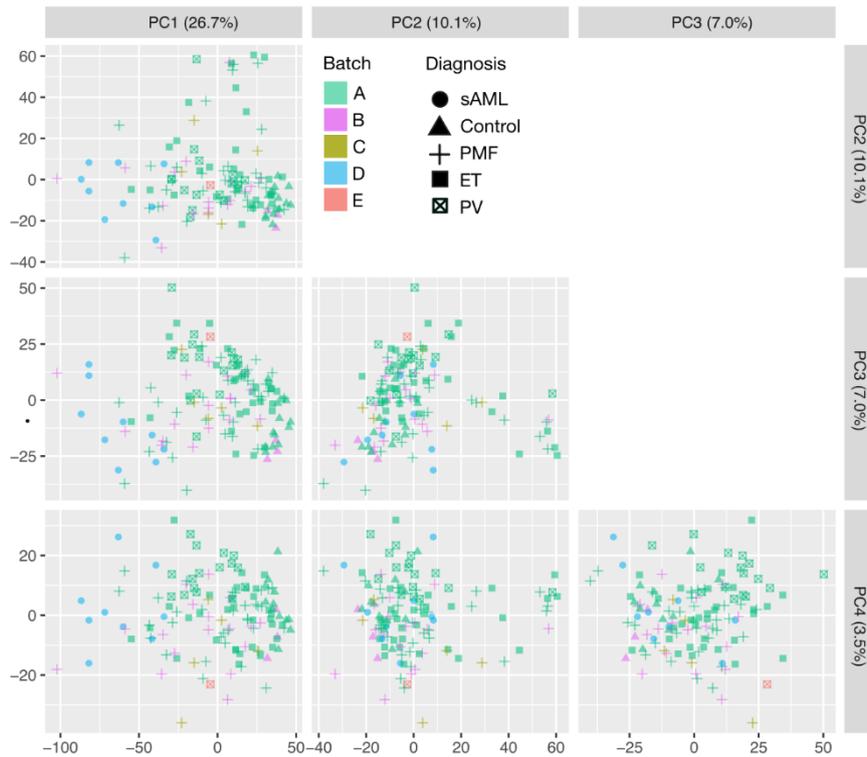
Figure 17: Principal component analysis of MPN patients. PCA analysis was performed on read counts of known and novel splice sites of 104 chronic MPN and 9 sAML patients. Only the 500 most variable splice sites were considered. Patients were colored by batch and symbols group patients by diagnosis.

We found overlaps of different disease entities (PV, ET, and PMF) of the chronic MPN patients whereas post-MPN sAML patients cluster distinctly from the rest of the chronic MPNs. *SF3B1* mutated patients reserved interesting positions within this biological continuum. Patients with *SF3B1*-K700E mutations showed splicing patterns which clustered them closer to secondary AML patients. Indeed, progression to post-MPN sAML was observed in patient P069 with a K700E mutation (**Figure 18**).
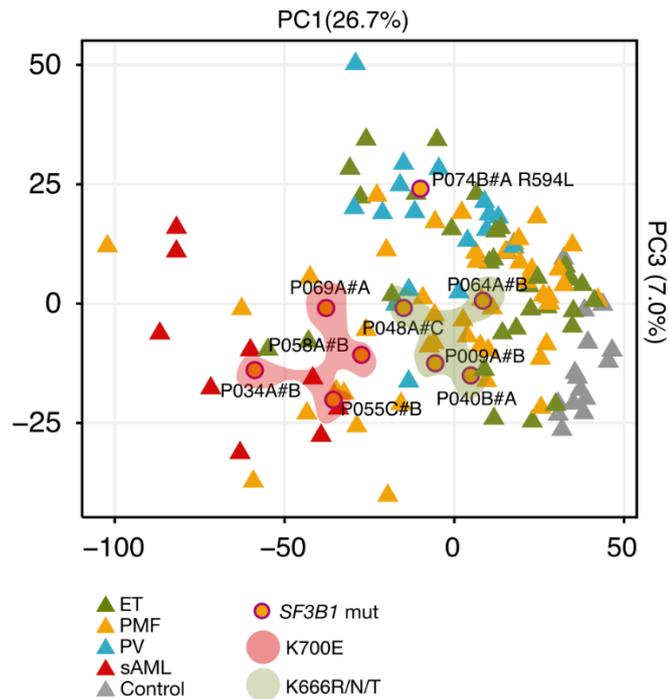
Figure 18: Principal component (PC) 1 and 3 shows distinct clustering of *SF3B1*-K700E mutated and secondary AML patients. PCA analysis was performed on read counts of known and novel splice sites of 104 chronic MPN and 9 sAML patients. Only the 500 most variable splice sites were considered. Patient samples were colored by diagnosis. *SF3B1* mutated patients are marked with circles. Red and green color shading label K700 and K666 mutations.

Second, if *SF3B1* hotspot mutations are associated with disease progression, we would expect worse overall outcome for these patients. Survival analysis including 18 *SF3B1* mutated and 47 *SF3B1* wildtype PMF patients showed a trend towards decreased overall survival (OS) (*P-value*=0.18) (**Figure 19A**). Although, part of this trend appears to be mediated through *CALR* mutated patients which exhibit a significant better OS compared to *JAK2* mutated patients (*P-value*=0.014) in our cohort (**Figure 19B**), we focused our survival analysis on *JAK2* (*SF3B1* co-mutated an unmutated) patients only, and could confirm -even with a low patient number (n=31)- a similar trend (*P-value*=0.25) (**Figure 19C**).
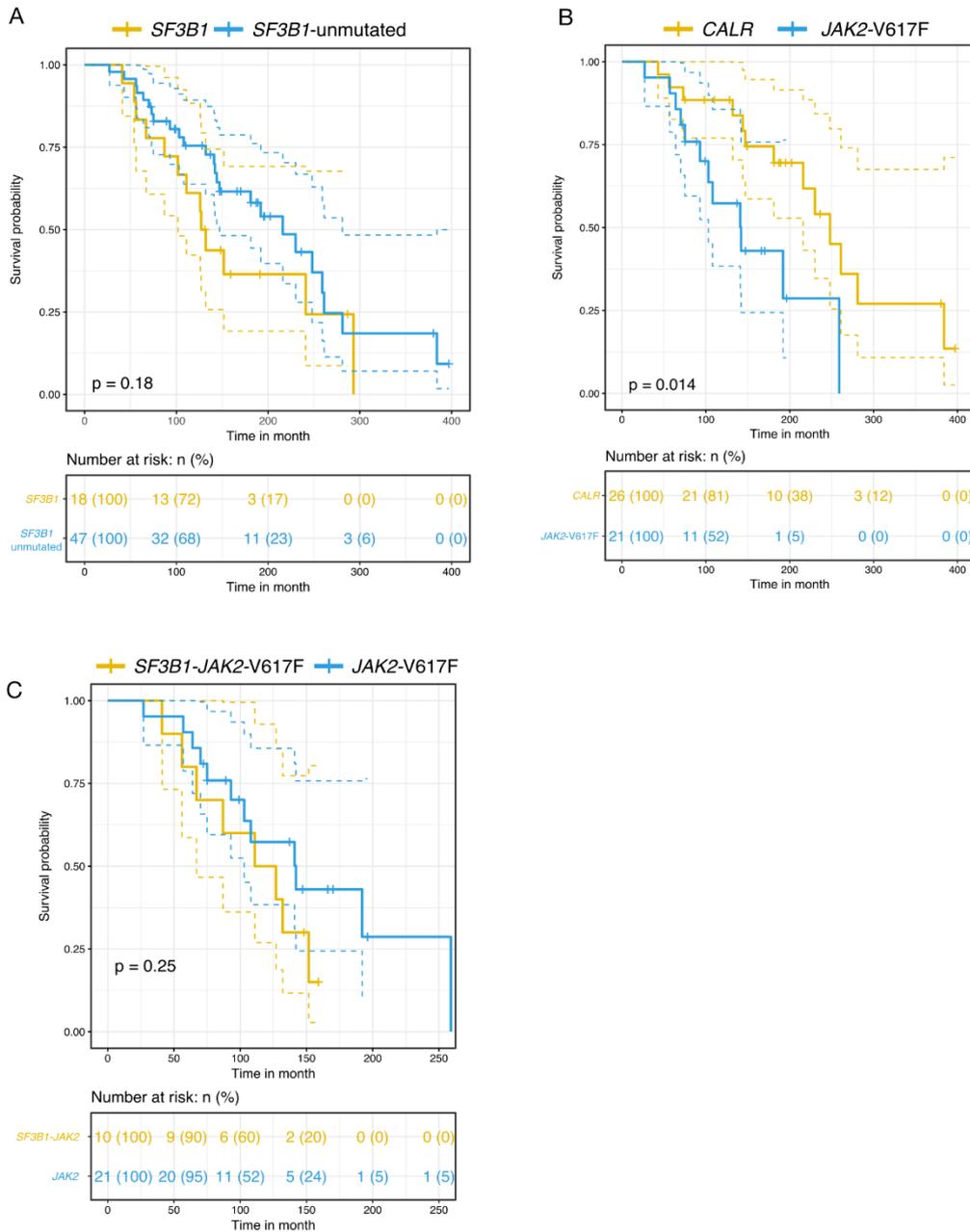
Figure 19: Kaplan-Meier survival analysis for *SF3B1* mutated PMF patients. Survival analysis was performed with R package *survival* and visualized with R package *survminer*. A log-rank or Mantel-Cox test was used to estimate *P-values*. *P-values* <0.05 were considered significant **(A)** Survival analysis of PMF patients only. PMF patients forming part of the transcriptome sequencing study and targeted sequencing study were pooled for this analysis (**Table 3**). **(B)** Survival analysis of PMF patients (only *SF3B1* unmutated) stratified by mutations in *CALR* (del52 or ins5) or *JAK2*-V617F **(C)** Survival analysis of PMF patients (only *JAK2*-V617F mutated) stratified by *SF3B1* co-mutation with *JAK2*-V617F.

4 DISCUSSION

To the best of our knowledge, this is the largest MPN centered effort to describe the mutational landscape of the transcriptome. This was accomplished through the versatile ways of processing and analyzing RNA-seq data. A similar study in size, analyzed the granulocyte transcriptional profile of 97 MPN patients (Rampal et al., 2014). However, this study used microarrays for gene expression profiling to derive signatures of activated pathways of *JAK2*, *CALR*, and *TET2* mutated patients. Although microarray technology is a suitable method for analyzing gene expression data and altered pathways, RNA-seq technology is superior in resolution and sensitivity. Additionally, an incredibly diverse set of workflows can be applied to RNA-seq data, making it feasible to answer distinct questions with a single dataset. The following paragraphs critically discuss the merits and caveats of these workflows and the biological insights we gained for the benefit of MPN patients.

Analysis of granulocyte clonality using RNA-seq provides putative novel biomarkers for granulocyte clonality for female MPN patients

We have implemented an approach to estimate clonality in granulocyte cells of MPN patients using RNA-seq. Szelinger et al. applied a similar workflow for the analysis of X-linked genetic conditions (Szelinger et al., 2014). However, they used a combined whole-exome and RNA-seq approach to estimate XCI ratios, which was not available for our cohort. Here we present a workflow that relies solely on RNA-seq data and report its first application on patients with hematological cancers. The strength of this method is the combined chromosome-wide usage of heterozygous variants (*global approach*) on the X chromosome instead of relying on single gene locus or combination of genes as clonality markers (*local approach*).

Reproducibility and robustness of the method was tested comparing XCI ratios of four technical replicates. A larger number of technical replicates is necessary to draw conclusive results, but a preliminary comparison of these samples showed high concordance between technical replicates. We did not benchmark our method of XCI ratio estimation with routine clinical assays such as DNA methylation-based assays as

HUMARA, which targets CAG short tandem repeats in the human androgen receptor (AR) gene (Allen et al., 1992). Although, it would be interesting to compare these two methods from a technical point of view, the results should be interpreted with caution. There is a considerable debate whether DNA methylation can reliably reflect the ratio of expression between the active and inactive X chromosome compared to quantitative allele expression methods (Busque et al., 2009; Swierczek et al., 2012).

Using this workflow, we compared single gene locus approaches used in transcription-based clonality assays to our global approach of XCI ratio estimation. We could rank genes based on strong or worse correlation (local versus global approach). *XIST* and *G6PD*, known genes involved in X-inactivation or clonality assays (Chen and Prchal, 2007), were among the highly ranked genes. However, only 9 and 23 females had heterozygous variants (or polymorphic markers) on the *XIST* and *G6PD* gene locus, respectively. This emphasizes the need for clonality assays inclusive for all females as exemplified with our global approach. We have also identified potentially new marker genes of clonality for X inactivation assays such as *CXorf65*, *WDR44* and *MED12* with 29, 16 and 29 informative females on these loci, respectively. None of these genes were listed as '*escape gene*' in a study examining escape of X-inactivation of genes across 23 tissues in humans (Slavney et al., 2015).

Skewing of X inactivation is frequently seen in healthy females (Amos-Landgraf et al., 2006). This represents a limiting factor for reliable interpretation of our results. However, this is an intrinsic problem shared by all clonality assays and can be controlled for with appropriate control tissue such as T cells, which was unfortunately not available for our cohort. Extreme skewing of more than 90:10 XCI-ratio is extremely rare in healthy females. It was not observed in any of our control females but in 33/58 cases of female MPN patients (6/9 PV, 8/20 ET, 19/29 PMF). Skewed lyonization in healthy females has been attributed to stochastic skewing due to a small pool of cells in which the inactivation initially occurs as well as stem cell depletion or selection forces (Busque et al., 1996; Chen and Prchal, 2007; Gale et al., 1997). Aging is also associated with an increase in clonal hematopoiesis. A series of publications have identified somatic mutations in genes such as *TET2*, *DNMT3A* in healthy, elderly individuals with clonal hematopoiesis (Jaiswal et al., 2014; Lindberg et al., 2014; Xie et al., 2014). However, using our RNA variant calling approach, none of our healthy females had mutations in these genes.

We think our approach to distinguish clonal from polyclonal patients could be of interest for triple-negative MPN patients where no driver gene has been identified yet. Our 3 triple-negative ET females had the lowest XCI ratios of all ET patients. Although the true nature of a polyclonal hematopoiesis would need to be confirmed with control tissue (analysis of T cell clonality) as previously explained, it is very likely that these patients have indeed a polyclonal hematopoiesis.

A study by Milosevic et al. determined the clonality status of 26 triple-negative PMF (n=3) and ET (n=23) female patients and available control tissue using HUMARA assays (Milosevic Feenstra et al., 2016). Of the 23 ET patients, 3 had a clonal, 9 had a polyclonal hematopoiesis, 7 were non-informative, and 4 cases were ambiguous.

First, the study reports 7 cases of non-informative patients, which represents 30% of triple-negative ET patients. This emphasizes the utility of our global approach of XCI ratio estimation, which is inclusive for all females.

Second, this study could show that polyclonal triple-negative MPN cases are common. Milosevic et al. hypothesized whether polyclonal triple-negative ET patients are essentially misclassified cases of hereditary MPN-like disorders with an unknown germline mutation responsible for disease pathogenesis. Hereditary thrombocytosis (HT) is a disease with shared clinical features with ET and without a known family history it is likely that HT cases could be misclassified as sporadic ET. HT patients have a polyclonal hematopoiesis, no disease progression, and a better disease prognosis than ET patients (Harutyunyan and Kralovics, 2012). It is therefore of essential importance to distinguish these two diseases in order to provide a better prognostic information to the patient and avoid unnecessary therapeutic intervention such as cytoreductive therapy. For cases with hereditary disorders strategies to identify germline mutations should be evaluated. For true triple-negative patients with a clonal hematopoiesis and no other somatic mutations, it would be worthwhile to extend whole-exome sequencing approaches to whole-genome for mutations in regulatory regions or RNA sequencing for fusion discovery to search for potential disease drivers.

Transcriptome-wide fusion discovery in MPN patients

We implemented a workflow for discovery and filtering of fusions using RNA-seq in a cohort of 104 MPN as well as 9 post-MPN sAML patients. A careful selection of fusion detection tools, filtering and sequencing strategy was necessary to ensure high quality prediction of fusion candidates. Given the small overlap of fusion detection tools, we

combined results from three fusion detection tools (*defuse*, *tophat-fusion*, and *soapfuse*), a practice recommended to overcome algorithm-specific biases and increase sensitivity (Abate et al., 2014; Liu et al., 2016). The union of all fusion detection tools resulted in a total of 13 non-recurrent unique fusion events in 12 MPN patients. We propose the following classification for our 13 fusion events:

Class A fusions *BCR-ABL1, XPO5-RUNX1*, and *KDM4B-CYHR1* were examples of fusion transcripts with intact exons where the genomic breakpoint most likely resides in intronic regions and the cDNA break point is located at the boundary of the exon. **Class B** fusions *FRY-MYCBP2*, *CUX1-intragenic*, *AHCTF1-SMYD3* and *INO80D-GPR1* had an intact exon of the 5' gene, but the downstream location of the cDNA breakpoint is either intronic, genomic or in the middle of an exon. Splice acceptor and donor sites are present in all these fusion genes. *NOX5-PNISR* is a fusion gene where both cDNA breakpoints are located within introns. The reciprocal transcript *PNISR-NOX5* was also detected. Finally, **class C** fusions *B2M-SELL*, *CSF3R-ZFP36*, *SAT1-MNDA*, *S100A8-PLEK* and *S100A9-CTSD*, were characterized by a stretch of 5-7 repetitive nucleotides within the cDNA breakpoint which are located within protein coding exons, UTR regions, or in one case, within intronic region. Additionally, no splice acceptor and/or donor sites are present.

Among the 13 fusions found in MPN and sAML patients, *BCR-ABL1* is the only known fusion. *BCR-ABL1* is a known hallmark fusion in chronic myeloid leukemia (CML). However, concurrent *JAK2*-V617F (Bornhäuser et al., 2007; Park et al., 2013) or *CALR* (Cabagnols et al., 2015a; Pagoni et al., 2014) mutations and *BCR-ABL1* translocations have been reported. In our cohort, the patient with a *BCR-ABL1* fusion was initially diagnosed with PMF in 2005 and treated with hydroxyurea and thromboreductin. At that timepoint (2005) no detectable *BCR-ABL1* fusion was found. Two years later the patient was tested positive for *BCR-ABL1* with a clone size of 25.9%. In 2008 patient blood was sampled and analyzed in 2014 using RNA-seq, which confirmed the existence of a *BCR-ABL1* fusion. Since February 2008, the patient has been treated with imatinib and thromboreductin. A follow-up in 2014 revealed a complete remission in respect to the *BCR-ABL1* clone but a persistent *JAK2* clone which is indicative for a clonal exchange in the patient. The detection of a known *BCR-ABL1* fusion in a PMF patient emphasizes the utility of RNA-seq based fusion discovery for diagnostic purposes.

A PMF patient with a *FRY-MYCBP2* fusion also presented with a large deletion on the large arm of chromosome 13, which was detected with SNP6.0 arrays. Low resolution of SNP6.0 array technology prohibited exact localization of genomic breakpoints, but the start and end of the aberration was located close to the *FRY* and *MYCBP2* gene locus. Therefore, the 13q deletion in this PMF patient likely leads to truncation of both fusion gene partners and formation of a fusion gene detectable on RNA level. The 13q deletion is a frequent aberration in MPN patients (Caramazza et al., 2011; Klampfl et al., 2011). Whether all 13q deletions lead to formation of *FRY-MYCBP2* fusion genes is unclear. 13q deletion can vary considerably in length, it is therefore unlikely that *FRY* or *MYCBP2* is a repetitive target of the 13q deletion.

Except for known and *in frame BCR-ABL1* oncogenic fusion, 10 fusions detected in chronic MPN were novel and *out of frame*. Although in the past, functional workup of candidate fusions has been focused on *in frame* fusions (Mertens et al., 2015), more recent efforts have identified gene truncations due to *out of frame* fusions as a mechanism to effectively inactivate one allele of the two genes at the fusion junction and likely resulting in haploinsufficiency of one or both of the rearranged genes (Duro et al., 1996; Rodriguez-Perales et al., 2015; Storlazzi et al., 2005). Therefore, we cannot exclude that any of the *out of frame* fusions follow similar mechanisms and are not simply passenger mutations.

Individual genes forming parts of MPN fusions have been previously implicated in tumorigenesis. *CUX1* is an example of a homeodomain-containing transcription factor which was shown to act as a haploinsufficient tumor suppressor. *CUX1* mutations have shown to provide engraftment advantage in human hematopoietic cells after transplantation into immune-deprived mice (McNerney et al., 2013; Wong et al., 2014). McNerney et al. reported a case in which *CUX1* was disrupted by a translocation fusing together exon 1 of *CUX1* with exon 2-4 of *CLDN7* (McNerney et al., 2013). The *CLDN7* exons are *out of frame* resulting in a premature stop codon. In our patient the *CUX1* fusion is mediated through an inversion event, fusing together exon 1 of *CUX1* gene with a genomic sequence and a stop codon is encountered after 99 nucleotides. In both events *CUX1* gene is disrupted after the first exon, most probably leading to a loss of function of *CUX1*. Other example of individuals genes with implication in cancer are *XPO5* (Melo et al., 2010; Shigeyasu et al., 2017), *RUNX2* (Li et al., 2016; Onodera et al., 2010), *MYCBP2* (Ge et al., 2015), *SMYD3* (Sarris et al., 2016), *CSF3R* (Plo et al., 2009), *B2M* (Pereira et al., 2017), and *KDM4B* (Qiu et al., 2015).

We did not observe any fusion events in 4 triple-negative ET patients. As previously reported, for 3 females ET patients we showed a likely polyclonal hematopoiesis and therefore a hereditary MPN-like disorders is a plausible diagnosis. It is therefore not surprising, that we did not find any somatic fusions, that drive the disease in these patients. Six fusions were found in PMF, 3 in ET and 2 in PV patients. However, given their low overall occurrence it was not possible to assign any statistically significant association in fusion frequency with disease subtypes. Two out of 9 post-MPN sAML had fusion genes (*INO80D-GPR1* and *XPO5-RUNX2*) not previously described. We did not detect *de novo* AML associated fusion genes in any of our 9 post-MPN sAML patients, emphasizing the distinct disease evolution of these two diseases (Milosevic et al., 2012).

We validated all fusions and mapped the cDNA breakpoint with sanger-sequencing on RNA level. Mapping the genomic breakpoint on DNA level using a primer combination approach was successful for *in frame* fusion *INO80D-GPR1*. The other fusions had large intronic sequences to span, which could likely be the cause why genomic breakpoint mapping failed for these fusions. We cannot exclude that some fusions with no genomic or additional supporting evidence could in fact represent trans-splicing events. Trans-splicing joins together exons from different genes, even if they are located on different chromosomes and generate a chimeric RNA without the existence of chromosomal aberrations (Akiva et al., 2006; Lei et al., 2016). These chimeric transcripts can be translated to proteins and can drive tumor development as has been shown for *JAZF1-JJAZ1*, a transcription mediated fusion located on different chromosomes (Li et al., 2008). In general, trans-splicing events are usually expressed at low levels compared to the native gene transcripts. We have shown that most of our fusion events found in MPN are indeed expressed at low level and would support the theory of trans-splicing event. However, this needs to be evaluated carefully, considering that except for *BCR-ABL1* and *INO80D-GPR1*, all reported fusions were *out of frame* and are therefore likely to be degraded if nonsense-mediated decay (NMD) rules apply. Therefore, a high expression of these transcripts is not expected.

We hypothesize that the formation of such a chimeric transcript on RNA level requires a proximity between both RNA molecules and consequently the involved genes of the fusion are likely to be situated in close vicinity within the 3-dimensional genome. It would be interesting to test this hypothesis using genome-wide chromosome conformation

capture (Hi-C) to identify if genes forming trans-splicing events are more likely to be close to each other compared to any two genes of the genome.

RNA-seq based variant calling and identification of high frequency mutations of SF3B1 in PMF patients

We implemented a workflow for SNV and Indel calling on RNA level using the STAR aligner (Engström et al., 2013) and GATK's best practice workflow for RNA variant calling. At the time of workflow development, there was no RNA-specific algorithm for somatic variant calling such a MuTect for DNA (Cibulskis et al., 2013). We therefore relied on GATK's Haplotype Caller optimized for DNA variant calling and diploid genomes. As a result, capturing low frequency mutations (<0.1 mutant allele frequency), was not feasible without prior knowledge of the mutation. The recent release of RNA-MuTect would offer the possibility to detect small subclones with low mutation frequency and would be an interesting supplement to the study of minimal residual disease in MPN (Yizhak et al., 2018). Additionally, for most samples there was no germline tissue available. We established a set of filters to compensate for the lack of control tissue. This filters included removal of common variants (annotated as common in publicly available databases), focusing on deleterious variants and extracting variants from a list of 86 genes frequently mutated in myeloid disease and MPN (Klampfl et al., 2013; Lundberg et al., 2014a; Nangalia et al., 2013).

Applying variant calling on transcriptome data, enabled us to interrogate the mutation status of frequently mutated genes in myeloid diseases. We found a high frequency (16.3%) of *SF3B1* mutated PMF patients in our RNA cohort. Although, *SF3B1* hotspot mutations are frequently found in MDS (20%) and CLL (18%) (Darman et al., 2015; Papaemmanuil et al., 2011), only few studies reported on *SF3B1* mutation frequency in MPN. Some reported no occurrence (Yoshida et al., 2011), others frequencies up to 6.5% in PMF (Lasho et al., 2012) (Nangalia et al., 2013).

To rule out the possibility that our initially determined *SF3B1* mutation frequency of 16.3% was due to a sampling bias, we sequenced an additional 113 PMF patients from our MPN biobank. We found the overall occurrence of *SF3B1* mutations in PMF to be 10.7%. Although, we could not reach the frequency found in our RNA cohort, a frequency of 10.7% strongly argues against *SF3B1* being an infrequent mutation in PMF patients.

Clinical relevance of SF3B1 mutations in PMF patients

Our re-sequencing efforts could show, that *SF3B1* mutations are not restricted to PMF patients. We found *SF3B1*-K666N in two PV patients. However, the frequency was very low (2/225 PV patients) and suggests that *SF3B1* hotspot mutations are associated with the more progressed form of MPN disease subtypes. Indeed, we observed evidence of disease progression in 6/18 (~33%) of *SF3B1* mutated patients. Additionally, we could show in a principal component analysis of read counts of splice sites, that patients with mutation type K700E but not K666 cluster close to sAML patients and distinctly from chronic MPN patients. Our clinical data shows progression in both mutation types (K700 and K666) and does not support the observation seen in the PCA analysis. Confounding factors (e.g. batch effects) restrict reliable interpretation of our results. Survival analysis showed a trend towards worse survival in *SF3B1* mutated patients. Different mutation types were pooled to increase statistical power. These results contradict previous survival analysis of *SF3B1* mutation in PMF patients, showing no influence in survival (Lasho et al., 2012). However, this study included less frequent and less described mutations H662 and N626 in their survival analysis, which could partly explain differing results.

Additional experiments and a larger patient cohort are required to show if *SF3B1* hotspot mutations are indeed progression associated.

SF3B1 hotspot mutations induce aberrant 3'splicing

Splicing defects caused by *SF3B1* mutations have been subject to extensive research especially to characterize the abnormal splicing patterns of mutated cells (Alsafadi et al., 2016; Bueno et al., 2016; Darman et al., 2015; DeBoever et al., 2015; Obeng et al., 2016; Pellagatti et al., 2018). Similar to other studies, we have identified cryptic 3' splicing as the most abundant splicing defect. However, the splicing patterns we described for PMF in this study are distinct from other cancers (even within other myeloid diseases). Therefore, this study is the first to report on MPN-specific splicing abnormalities caused by *SF3B1* mutants. It is unclear at this point what factors may be responsible for the minimal overlap of splicing abnormalities among cancers. Differences either in the cell types compared and/or the specific mutagenic landscapes of cancers can be such factors. A small number of splicing defects were shared among all cancers or at least found in 6 of 7. Among the 3' splice sites with the highest impact on predicted amino acid change, the following affected genes had a high overlap with other cancer types: *MAP3K7, OX1AL, RWDD4, CC2D1, PRPF38A*. Interestingly, a recent publication on splicing

defects, identified mis-spliced *MAP3K7* in *SF3B1* as a driver of hyperactive NF-κB signaling (Lee et al., 2018).

Whether mis-spliced *MAP3K7* or any other aberrantly spliced genes identified in the study might functionally collaborate with the growth promoting effect of *JAK2*-V617F or *CALR*-mutated cells, requires further investigation. A different study by Kim et al., 2015 (Kim et al., 2015b) could propose a mechanistic relationship between mutations in splicing machinery gene *SRSF2* and miss-spliced and deficient *EZH2* gene – a key hematopoietic regulator. None of our *SF3B1* mutated patients showed any splicing defects in *EZH2* or lower mRNA level. Therefore, it remains to be seen if among *SF3B1* affected genes some might act as tumor suppressors and increase the fitness of the MPN disease clone.

Discovery of putative neoantigens in MPN patients

We established a workflow for systematic discovery of putative neoantigens in MPN patients. This workflow involved translation of mutation into proteins, extracting MHC-I genotype information for each patient, predicting binding affinities with NetMHCpan for all 8, 9 and 10mer peptides for each patient individually, taking into consideration all mutations identified and the patient's personal MHC-I molecules.

Although the number of Indels and SNV identified per patient is relatively low as expected for MPN patients (Nangalia et al., 2013), we could still identify 149 unique neoantigens, with an occurrence of 61.9% in MPN patients that could have potential use for personalized cancer immunotherapy approaches. Of these 149 unique neoantigens, 64 were SNV and 85 were derived from frameshift mutations. A total of 27 unique neoantigens from 85 frameshift mutations were derived mostly from *CALR*, followed by *ASXL*1 and *TET2*. Interestingly, we identified a unique property of the *CALR* neoantigens of binding several common MHC-I molecules. Consequently, 75% of *CALR*-mutated patients have a *CALR* neoantigen predicted to be presented by their personal MHC-I molecules. We also identified 16 unique neoantigens for *MPL*-W515K/L/A mutations for 5 of the most common HLA haplotypes.

In addition, we identified 169 peptides derived from aberrantly spliced genes in *SF3B1* mutated patients with a predicted binding to at least 1 MHC-I molecule. The value and pitfalls of using neoantigens from mis-spliced genes in *SF3B1* mutated patients are described in the discussion of results **section 3.2.**

In summary, we believe that our results may serve as a framework for systematic mining of neoantigens from a diverse set of mutation classes using RNA-seq. The discovered putative neoantigens may serve as a resource for generation of individualized vaccines or adoptive cell-based therapies in MPN.

5 MATERIALS AND METHODS

5.1 Materials and Methods for results section 3.1

5.1.1 Data processing and filtering

Patients were selected, sequenced and RNA-seq data was processed and single nucleotide variants (SNVs) were called as described in **section 3.2 (Material and Methods)**. SNV calls and variant allele frequencies (VAFs) from the X chromosome were extracted from patients and healthy controls and annotated with ANNOVAR (release 2015-06-17) (Wang et al., 2010). SNVs were filtered as follows (**Figure 20**):
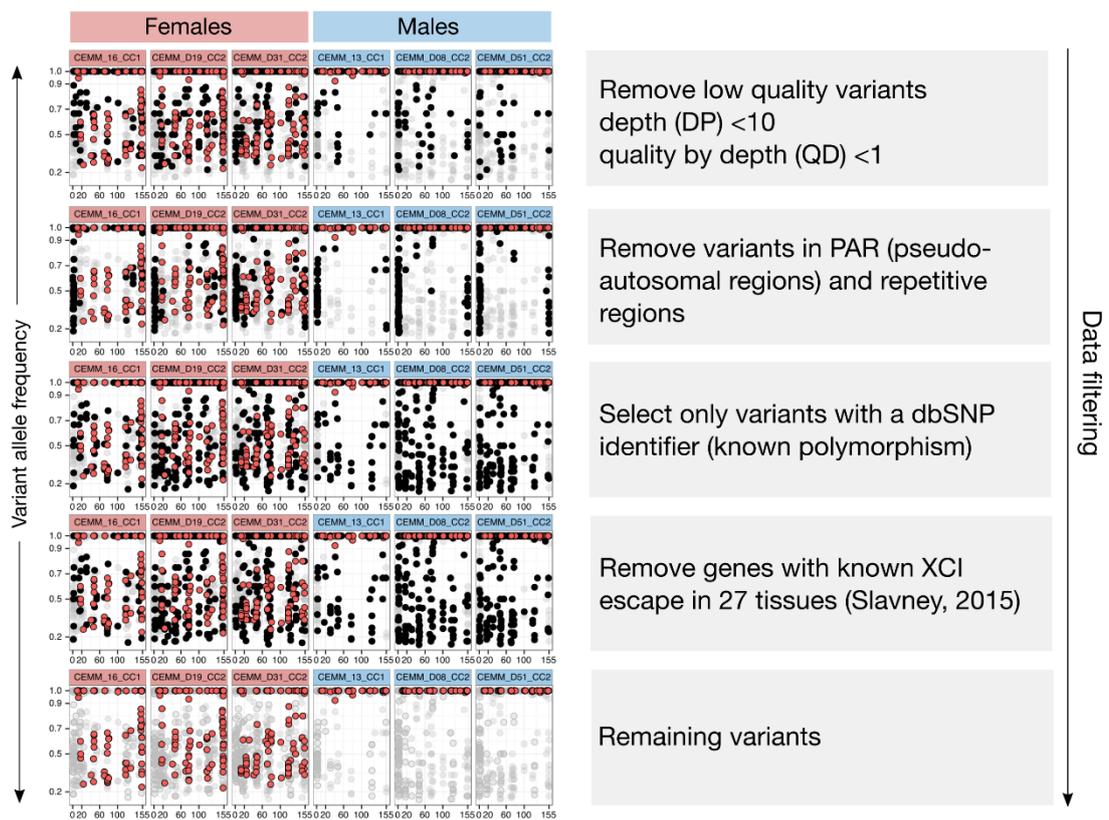
Figure 20: SNVs calls extracted from the X chromosome for 3 representative females and males. The graph on the left shows on the x-axis the chromosomal coordinates of the X chromosome (from 0-155kb). Each dot represents an SNVs. Black and grey dots are filtered SNVs, red dots remain after all filtering steps were applied.

SNVs with a quality by depth (QD) less than 1 and less than 10 reads covering the variant were removed. Variant calls within segmental duplications (g*enomicSuperDups;*

hgdownload.soe.ucsc.edu/goldenPath/hg19/database/genomicSuperDups.txt.gz) larger than 1kB and more than 90% similarity were excluded (these included pseudo-autosomal regions PAR1 & PAR2 of the X and Y chromosome). Additionally, variants overlapping low complexity regions as well as repetitive regions defined by *RepeatMasker* (http://hgdownload.cse. ucsc.edu/goldenPath/hg19/database/rmsk.txt.gz) were excluded from further analysis. Only SNVs reported as polymorphisms (SNPs) according to dbSNP (version 142) were kept. Genes reported as X chromosome escape genes in 27 tissues were removed (Slavney et al., 2015).

5.1.2 Estimating X chromosome inactivation ratio (XCI ratio)

The XCI ratio was calculated for each sample separately. For each sample, variant allele frequencies (VAF) of heterozygous variants (VAF <0.97) after filtering were considered. Phasing information was not available for our samples, we therefore could not distinguish whether a variant has a maternal or paternal expression (**Figure 21**).
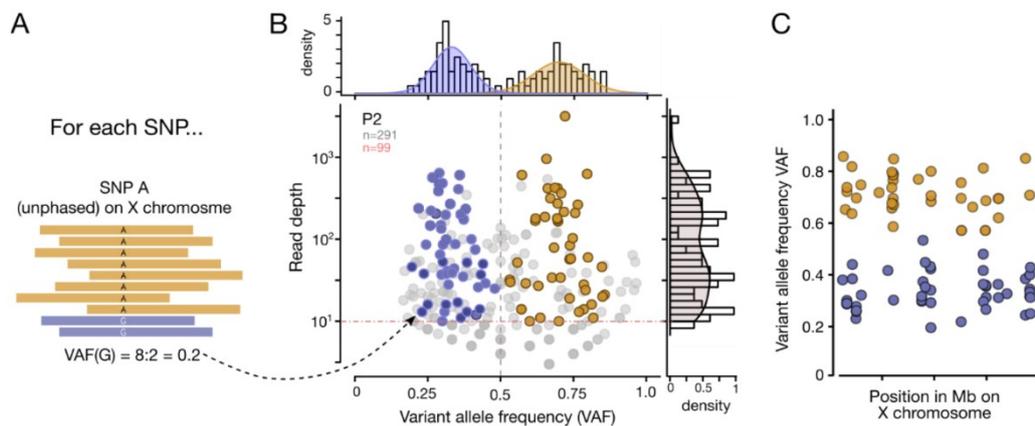
Figure 21: Variant allele frequency (VAF) was estimated for each SNV and for each patient. (A) For each SNV or SNP (single nucleotide polymorphism), the variant allele frequency was extracted. **(B)** For each sample the extracted SNPs can be plotted as a function of variant allele frequency and read depth. For sample P2, 291 SNV were extracted and 99 variants remained after filtering. The SNVs show a biomodal distribution with variants expressed from the maternal or paternal allele. However, phasing information is not known, therefore a distinction between maternal or paternal expressed variant is not possible. **(C)** Variant allele frequencies for each SNV plotted along their respective position on the X chromosome.

In samples where the maternal and paternal variants are equally expressed within a pool of cells, variant allele frequencies exhibit a unimodal distribution (assessed using Hartigan's dip test (*diptest*, R package) and for these samples the XCI ratio was estimated as the mean of variant allele frequencies extracted for each sample (**Figure 22**). A bias in the expression of maternal and paternal alleles, may shift the allele expression to a

bimodal distribution, for these cases a mixture model was fit using R package *bayesmix* to estimate the parameters of each component of the distribution. The calculated mean for each sample is an indirect measurement of the expression status of each inherited (maternal or paternal) chromosome. For samples with less than 3 heterozygous variants no statistic was calculated, and a complete skew of the sample was assumed. Consequently, an XCI ratio of 1 (= 100) was assigned for that sample.
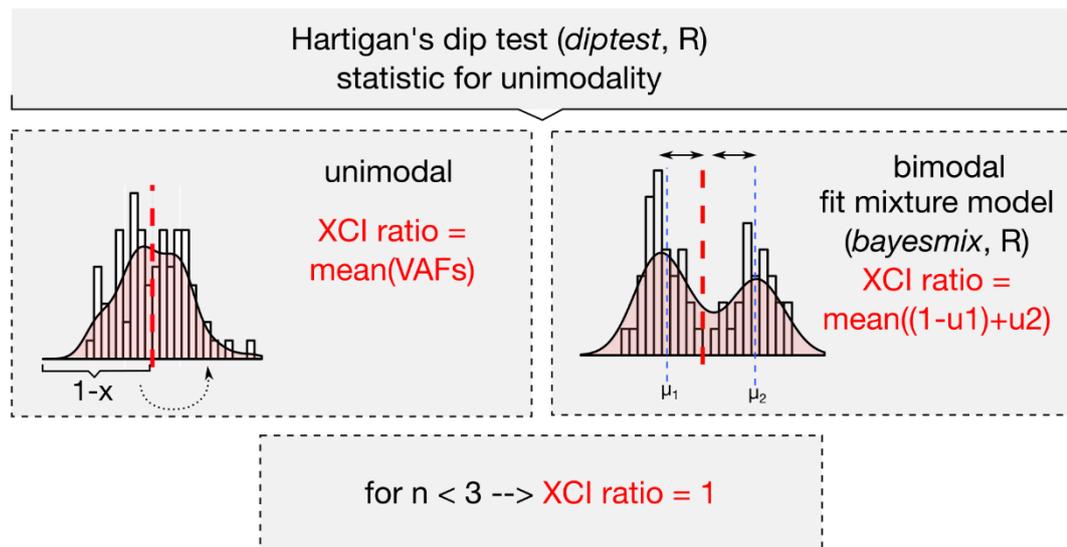
Figure 22: Calculation of X-chromosome inactivation ratio for each sample. Hardigan's dip test (*diptest*, R) was used to for unimodality for each sample. Depending on the outcome, if a sample had a unimodal distribution, the XCI ratio was estimated as the mean of the variant allele frequencies (VAFs). For bimodal distributions a mixture model was fit to the data to estimate u1 and u2 parameters for each component of the distribution. For samples with less than 3 datapoints the XCI ratio was assigned to 1 (=100).

Using this method, the XCI ratio can range from 50:50 (random) to 100:0 (complete skew). Based on previous observation and publications, the skew of X chromosome inactivation is assigned as random from 50:50 – 75:25, a moderate skew is between 75:25 – 90:10, and an extreme skew is above 90:10 (Szelinger et al., 2014) (**Figure 23**).
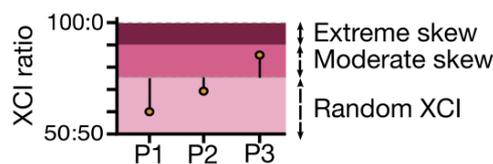
Figure 23: X chromosome inactivation skew. Szelinger et al. proposed based on previous observation the following classification: Samples (P1-P3) with an XCI ratio between 50:50 to 75:25 are considered to have a random X inactivation, from 75:25 to 90:10 a moderate skew and everything above 90:10 is considered an extreme skew.

5.2 Material and Methods for results section 3.3

5.2.1 Targeted resequencing

Samples were processed using the TruSight Myeloid Sequencing Panel which allowed for quantitative evaluation of mutations in 54 genes implicated in myeloid diseases. In addition, we generated and sequenced a single amplicon covering two *SF3B1* hotspot mutation sites (K666 and K700). PCR was performed using Herculase Fusion Polymerase (Agilent, Santa Clara, CA) applying standard conditions as recommended by the manufacturer at 55° annealing temperature, using primer pools (**Table 4**).

Table 4: Primer pool sequence.

Primer name	Primer sequence
SF3B1_K666_K700_F0	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-GAGTTGCTGCTTCAGCCAAG
SF3B1_K666_K700_F1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNGAGTTGCTGCTTCAGCCAAG
SF3B1_K666_K700_F2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNGAGTTGCTGCTTCAGCCAAG
SF3B1_K666_K700_F3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNGAGTTGCTGCTTCAGCCAAG
SF3B1_K666_K700_R0	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG(staggers)GCAAAAGCAAGAAGTCCTGG
SF3B1_K666_K700_R1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNGCAAAAGCAAGAAGTCCTGG
SF3B1_K666_K700_R2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNGCAAAAGCAAGAAGTCCTGG
SF3B1_K666_K700_R3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNGCAAAAGCAAGAAGTCCTGG

Primers included staggers required for amplicon diversity and resulting amplicons comprised primer binding sites required for sample indexing using the Nextera XT Index Kit v2 (Sets A-D, 384 indexes; Illumina San Diego, CA). Primers were pooled equimolar and used at standard concentrations.

DNA isolated from granulocytes or peripheral blood mononuclear cells (PBMCs) was processed according to the manufacturer's instructions to generate indexed amplicon-based libraries. Equimolar amounts of libraries were pooled into multiplexes (96-plexes for TruSight, 384-plexes for the single-amplicon approach) which were then sequenced 150bp paired-end on an Illumina HiSeq3000 or MiSeq instruments for the TruSight Myeloid Panel and the single-amplicon approach, respectively. Read alignment and variant calling was performed using the BaseSpace software (Illumina, San Diego, CA). Variants called in transcribed regions or at splice sites were selected and further filtered for common variation as described in results **section 3.2** *Variant calling on RNA sequencing data*. Other filters were adjusted for TruSight targeted sequencing and included insufficient sequencing read depth (<200) and low variant allele frequency (VAF <0.05).

6 REFERENCES

Abate, F., Acquaviva, A., Paciello, G., Foti, C., Ficarra, E., Ferrarini, A., Delledonne, M., Iacobucci, I., Soverini, S., Martinelli, G., et al., 2012. Bellerophontes: an RNA-Seq data analysis framework for chimeric transcripts discovery based on accurate fusion model. Bioinformatics 28, 2114–21.

Abate, F., Zairis, S., Ficarra, E., Acquaviva, A., Wiggins, C.H., Frattini, V., Lasorella, A., Iavarone, A., Inghirami, G., Rabadan, R., 2014. Pegasus: a comprehensive annotation and prediction tool for detection of driver gene fusions in cancer. BMC Syst. Biol. 8, 97.

Abdel-Wahab, O., Pardanani, A., Rampal, R., Lasho, T.L., Levine, R.L., Tefferi, A., 2011. DNMT3A mutational analysis in primary myelofibrosis, chronic myelomonocytic leukemia and advanced phases of myeloproliferative neoplasms. Leukemia 25, 1219–1220.

Abdulkarim, K., Girodon, F., Johansson, P., Maynadié, M., Kutti, J., Carli, P.-M., Bovet, E., Andréasson, B., 2009. AML transformation in 56 patients with Ph− MPD in two well defined populations. Eur. J. Haematol. 82, 106–111.

Adamson, J.W., Fialkow, P.J., Murphy, S., Prchal, J.F., Steinmann, L., 1976. Polycythemia Vera: Stem-Cell and Probable Clonal Origin of the Disease. N. Engl. J. Med. 295, 913–916.

Akada, H., Yan, D., Zou, H., Fiering, S., Hutchison, R.E., Mohi, M.G., 2010. Conditional expression of heterozygous or homozygous Jak2V617F from its endogenous promoter induces a polycythemia vera-like disease. Blood 115, 3589–97.

Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A., Sorek, R., 2006. Transcription-mediated gene fusion in the human genome. Genome Res 16, 30–36.

Alkan, C., Coe, B.P., Eichler, E.E., 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12, 363–76.

Allen, R.C., Zoghbi, H.Y., Moseley, A.B., Rosenblatt, H.M., Belmont, J.W., 1992. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am. J. Hum. Genet. 51, 1229–39.

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., et al., 2016. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. 7, 10615.

Amos-Landgraf, J.M., Cottle, A., Plenge, R.M., Friez, M., Schwartz, C.E., Longshore, J., Willard, H.F., 2006. X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. Am. J. Hum. Genet. 79, 493–9.

Araki, M., Yang, Y., Masubuchi, N., Hironaka, Y., Takei, H., Morishita, S., Mizukami,

Y., Kan, S., Shirane, S., Edahiro, Y., et al., 2016. Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms. Blood 127.

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., Vardiman, J.W., 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127, 2391–2405.

Asmann, Y.W., Hossain, A., Necela, B.M., Middha, S., Kalari, K.R., Sun, Z., Chai, H.-S., Williamson, D.W., Radisky, D., Schroth, G.P., et al., 2011. A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. Nucleic Acids Res. 39, e100.

Augui, S., Nora, E.P., Heard, E., 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. 12, 429–442.

Ballen, K.K., Shrestha, S., Sobocinski, K.A., Zhang, M.-J., Bashey, A., Bolwell, B.J., Cervantes, F., Devine, S.M., Gale, R.P., Gupta, V., et al., 2010. Outcome of Transplantation for Myelofibrosis. Biol. Blood Marrow Transplant. 16, 358–367.

Balligand, T., Achouri, Y., Pecquet, C., Chachoua, I., Nivarthi, H., Marty, C., Vainchenker, W., Plo, I., Kralovics, R., Constantinescu, S.N., 2016. Pathologic activation of thrombopoietin receptor and JAK2-STAT5 pathway by frameshift mutants of mouse calreticulin. Leukemia 30, 1775–1778.

Barbui, T., Vannucchi, A.M., Buxhofer-Ausch, V., De Stefano, V., Betti, S., Rambaldi, A., Rumi, E., Ruggeri, M., Rodeghiero, F., Randi, M.L., et al., 2015. Practice-relevant revision of IPSET-thrombosis based on 1019 patients with WHO-defined essential thrombocythemia. Blood Cancer J. 5, e369.

Barosi, G., Viarengo, G., Pecci, A., Rosti, V., Piaggio, G., Marchetti, M., Frassoni, F., 2001. Diagnostic and clinical relevance of the number of circulating CD34(+) cells in myelofibrosis with myeloid metaplasia. Blood 98, 3249–55.

Baxter, E.J., Scott, L.M., Campbell, P.J., East, C., Fourouclas, N., Swanton, S., Vassiliou, G.S., Bench, A.J., Boyd, E.M., Curtin, N., et al., 2005. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet (London, England) 365, 1054–61.

Beer, P.A., Campbell, P.J., Scott, L.M., Bench, A.J., Erber, W.N., Bareford, D., Wilkins, B.S., Reilly, J.T., Hasselbalch, H.C., Bowman, R., et al., 2008. MPL mutations in myeloproliferative disorders: analysis of the PT-1 cohort. Blood 112, 141–149.

Bellanne-Chantelot, C., Chaumarel, I., Labopin, M., Bellanger, F., Barbu, V., De Toma, C., Delhommeau, F., Casadevall, N., Vainchenker, W., Thomas, G., et al., 2006. Genetic and clinical implications of the Val617Phe JAK2 mutation in 72 families with myeloproliferative disorders. Blood 108, 346–352.

Benelli, M., Pescucci, C., Marseglia, G., Severgnini, M., Torricelli, F., Magi, A., 2012. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. Bioinformatics 28, 3232–9.

Berletch, J.B., Yang, F., Xu, J., Carrel, L., Disteche, C.M., 2011. Genes that escape from X inactivation. Hum. Genet. 130, 237–45.

Besses, C., Álvarez-Larrán, A., Martínez-Avilés, L., Mojal, S., Longarón, R., Salar, A.,

Florensa, L., Serrano, S., Bellosillo, B., 2011. Modulation of JAK2 V617F allele burden dynamics by hydroxycarbamide in polycythaemia vera and essential thrombocythaemia patients. Br. J. Haematol. 152, 413–419.

Beutler, E., Collins, Z., Irwin, L.E., 1967. Value of Genetic Variants of Glucose-6-Phosphate Dehydrogenase in Tracing the Origin of Malignant Tumors. N. Engl. J. Med. 276, 389–391.

Beutler, E., Yeh, M., Fairbanks, V.F., 1962. The normal human female as a mosaic of X-chromosome activity: studies using the gene for C-6-PD-deficiency as a marker. Proc. Natl. Acad. Sci. U. S. A. 48, 9–16.

Biankin, A. V, Waddell, N., Kassahn, K.S., Gingras, M.-C., Muthuswamy, L.B., Johns, A.L., Miller, D.K., Wilson, P.J., Patch, A.-M., Wu, J., et al., 2012. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491, 399–405.

Bonicelli, G., Abdulkarim, K., Mounier, M., Johansson, P., Rossi, C., Jooste, V., Andreasson, B., Maynadié, M., Girodon, F., 2013. Leucocytosis and thrombosis at diagnosis are associated with poor survival in polycythaemia vera: a population-based study of 327 patients. Br. J. Haematol. 160, 251–254.

Bornhäuser, M., Mohr, B., Oelschlaegel, U., Bornhäuser, P., Jacki, S., Ehninger, G., Thiede, C., 2007. Concurrent JAK2(V617F) mutation and BCR-ABL translocation within committed myeloid progenitors in myelofibrosis. Leukemia 21, 1824–1826.

Bousquet, M., Brousset, P., 2006. Myeloproliferative disorders carrying the t(8;9) (PCM1-JAK2) translocation. Hum. Pathol. 37, 500–500.

Briere, J., El-Kassar, N., 1998. Clonality markers in polycythaemia and primary thrombocythaemia. Baillière's Clin. Haematol. 11, 787–801.

Brown, S.D., Warren, R.L., Gibb, E.A., Martin, S.D., Spinelli, J.J., Nelson, B.H., Holt, R.A., 2014. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res. 24, 743–50.

Bueno, R., Stawiski, E.W., Goldstein, L.D., Durinck, S., De Rienzo, A., Modrusan, Z., Gnad, F., Nguyen, T.T., Jaiswal, B.S., Chirieac, L.R., et al., 2016. Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. Nat. Genet. 48, 407–416.

Busque, L., Mio, R., Mattioli, J., Brais, E., Blais, N., Lalonde, Y., Maragh, M., Gilliland, D.G., 1996. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. Blood 88, 59–65.

Busque, L., Paquette, Y., Provost, S., Roy, D.-C., Levine, R.L., Mollica, L., Gary Gilliland, D., 2009. Skewing of X-inactivation ratios in blood cells of aging women is confirmed by independent methodologies. Blood 113, 3472–3474.

Busque, L., Patel, J.P., Figueroa, M.E., Vasanthakumar, A., Provost, S., Hamilou, Z., Mollica, L., Li, J., Viale, A., Heguy, A., et al., 2012. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nat. Genet. 44, 1179–81.

Cabagnols, X., Defour, J.P., Ugo, V., Ianotto, J.C., Mossuz, P., Mondet, J., Girodon, F., Alexandre, J.H., Mansier, O., Viallard, J.F., et al., 2015a. Differential association of calreticulin type 1 and type 2 mutations with myelofibrosis and essential thrombocytemia: relevance for disease evolution. Leukemia 29, 249–252.

Cabagnols, X., Favale, F., Pasquier, F., Messaoudi, K., Defour, J.P., Ianotto, J.C., Marzac, C., Le Couédic, J.P., Droin, N., Chachoua, I., et al., 2015b. Presence of atypical thrombopoietin receptor (MPL) mutations in triple negative essential thrombocythemia patients. Blood 127, 333–342.

Campbell, P.J., Green, A.R., 2006. The myeloproliferative disorders. N. Engl. J. Med. 355, 2452–2466.

Caramazza, D., Begna, K.H., Gangat, N., Vaidya, R., Siragusa, S., Van Dyke, D.L., Hanson, C., Pardanani, A., Tefferi, A., 2011. Refined cytogenetic-risk categorization for overall and leukemia-free survival in primary myelofibrosis: a single center study of 433 patients. Leukemia 25, 82–88.

Carrara, M., Beccuti, M., Lazzarato, F., Cavallo, F., Cordero, F., Donatelli, S., Calogero, R.A., 2013. State-of-the-Art Fusion-Finder Algorithms Sensitivity and Specificity. Biomed Res Int 2013, 340620.

Carrel, L., Willard, H.F., 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434, 400–404.

Carreno, B.M., Magrini, V., Becker-Hapak, M., Kaabinejadian, S., Hundal, J., Petti, A.A., Ly, A., Lie, W.-R., Hildebrand, W.H., Mardis, E.R., et al., 2015. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. Science 348, 803–8.

Cerquozzi, S., Tefferi, A., 2015. Blast transformation and fibrotic progression in polycythemia vera and essential thrombocythemia: a literature review of incidence and risk factors. Blood Cancer J. 5, e366.

Cha, E., Klinger, M., Hou, Y., Cummings, C., Ribas, A., Faham, M., Fong, L., 2014. Improved Survival with T Cell Clonotype Stability After Anti-CTLA-4 Treatment in Cancer Patients. Sci. Transl. Med. 6, 238ra70-238ra70.

Chachoua, I., Pecquet, C., El-Khoury, M., Nivarthi, H., Albu, R.-I., Marty, C., Gryshkova, V., Defour, J.-P., Vertenoeil, G., Ngo, A., et al., 2016. Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. Blood 127, 1325–1335.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al., 2012. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat. Genet. 44, 23–31.

Chen, G.L., Prchal, J.T., 2007. X-linked clonality testing: interpretation and limitations. Blood 110, 1411–9.

Chen, K., Wallis, J.W., Kandoth, C., Kalicki-Veizer, J.M., Mungall, K.L., Mungall, A.J., Jones, S.J., Marra, M.A., Ley, T.J., Mardis, E.R., et al., 2012. BreakFusion: targeted assembly-based identification of gene fusions in whole transcriptome paired-end sequencing data. Bioinformatics 28, 1923–1924.

Chu, H.-T., Hsiao, W.W.L., Chen, J.-C., Yeh, T.-J., Tsai, M.-H., Lin, H., Liu, Y.-W., Lee, S.-A., Chen, C.-C., Tsao, T.T.H., et al., 2013. EBARDenovo: highly accurate de novo assembly of RNA-Seq with efficient chimera-detection. Bioinformatics 29, 1004–1010.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., Getz, G., 2013. Sensitive detection of

somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. 31, 213–9.

Cieślik, M., Chinnaiyan, A.M., 2017. Cancer transcriptome profiling at the juncture of clinical translation. Nat. Rev. Genet.

Curnutte, J.T., Hopkins, P.J., Kuhl, W., Beutler, E., 1992. Studying X inactivation. Lancet (London, England) 339, 749.

Dameshek, W., 1951. Some speculations on the myeloproliferative syndromes. Blood 6, 372–375.

Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., et al., 2015. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. Cell Rep. 13, 1033–45.

DeBoever, C., Ghia, E.M., Shepard, P.J., Rassenti, L., Barrett, C.L., Jepsen, K., Jamieson, C.H.M., Carson, D., Kipps, T.J., Frazer, K.A., 2015. Transcriptome Sequencing Reveals Potential Mechanism of Cryptic 3' Splice Site Selection in SF3B1-mutated Cancers. PLOS Comput. Biol. 11, e1004105.

Defour, J.-P., Chachoua, I., Pecquet, C., Constantinescu, S.N., 2016. Oncogenic activation of MPL/thrombopoietin receptor by 17 mutations at W515: implications for myeloproliferative neoplasms. Leukemia 30, 1214–1216.

Deininger, M., Radich, J., Burn, T.C., Huber, R., Paranagama, D., Verstovsek, S., 2015. The effect of long-term ruxolitinib treatment on JAK2p.V617F allele burden in patients with myelofibrosis. Blood 126, 1551–1554.

Deisseroth, A., Kaminskas, E., Grillo, J., Chen, W., Saber, H., Lu, H.L., Rothmann, M.D., Brar, S., Wang, J., Garnett, C., et al., 2012. U.S. Food and Drug Administration Approval: Ruxolitinib for the Treatment of Patients with Intermediate and High-Risk Myelofibrosis. Clin. Cancer Res. 18, 3212–3217.

Delhommeau, F., Dupont, S., Valle, V. Della, James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.-P., Robert, F., Alberdi, A., et al., 2009. Mutation in *TET2* in Myeloid Cancers. N. Engl. J. Med. 360, 2289–2301.

Ding, J., Komatsu, H., Wakita, A., Kato-Uranishi, M., Ito, M., Satoh, A., Tsuboi, K., Nitta, M., Miyazaki, H., Iida, S., et al., 2004. Familial essential thrombocythemia associated with a dominant-positive activating mutation of the c-MPL gene, which encodes for the receptor for thrombopoietin. Blood 103, 4198–4200.

Dudley, M.E., Wunderlich, J.R., Robbins, P.F., Yang, J.C., Hwu, P., Schwartzentruber, D.J., Topalian, S.L., Sherry, R., Restifo, N.P., Hubicki, A.M., et al., 2002. Cancer Regression and Autoimmunity in Patients After Clonal Repopulation with Antitumor Lymphocytes. Science (80-.). 298, 850–854.

DuPage, M., Mazumdar, C., Schmidt, L.M., Cheung, A.F., Jacks, T., 2012. Expression of tumour-specific antigens underlies cancer immunoediting. Nature 482, 405–409.

Duro, D., Bernard, O., Della Valle, V., Leblanc, T., Berger, R., Larsen, C.J., 1996. Inactivation of the P16INK4/MTS1 gene by a chromosome translocation t(9;14)(p21-22;q11) in an acute lymphoblastic leukemia of B-cell type. Cancer Res. 56, 848–54.

Dvinge, H., Kim, E., Abdel-Wahab, O., Bradley, R.K., 2016. RNA splicing factors as oncoproteins and tumour suppressors. Nat. Rev. Cancer 16, 413–430.

Eder-Azanza, L., Navarro, D., Aranaz, P., Novo, F.J., Cross, N.C.P., Vizmanos, J.L., 2014. Bioinformatic analyses of CALR mutations in myeloproliferative neoplasms support a role in signaling. Leukemia 28, 2106–2109.

el-Kassar, N., Hetet, G., Brière, J., Grandchamp, B., 1997. Clonality analysis of hematopoiesis in essential thrombocythemia: advantages of studying T lymphocytes and platelets. Blood 89, 128–34.

Elf, S., Abdelfattah, N.S., Chen, E., Perales-Patón, J., Rosen, E.A., Ko, A., Peisker, F., Florescu, N., Giannini, S., Wolach, O., et al., 2016. Mutant Calreticulin Requires Both Its Mutant C-terminus and the Thrombopoietin Receptor for Oncogenic Transformation. Cancer Discov. 6, 368–81.

Elliott, M.A., Tefferi, A., 2005. Thrombosis and haemorrhage in polycythaemia vera and essential thrombocythaemia. Br. J. Haematol. 128, 275–290.

Engström, P.G., Steijger, T., Sipos, B., Grant, G.R., Kahles, A., Rätsch, G., Goldman, N., Hubbard, T.J., Harrow, J., Guigó, R., et al., 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat. Methods 10, 1185–1191.

Fialkow, P., Faguet, G., Jacobson, R., Vaidya, K., Murphy, S., 1981. Evidence that essential thrombocythemia is a clonal disorder with origin in a multipotent stem cell. Blood 58.

Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., et al., 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 39, D945-50.

Francis, R.W., Thompson-Wicking, K., Carter, K.W., Anderson, D., Kees, U.R., Beesley, A.H., 2012. Fusionfinder: A software tool to identify expressed gene fusion candidates from RNA-seq data. PLoS One 7.

Furney, S.J., Pedersen, M., Gentien, D., Dumont, A.G., Rapinat, A., Desjardins, L., Turajlic, S., Piperno-Neumann, S., de la Grange, P., Roman-Roman, S., et al., 2013. SF3B1 mutations are associated with alternative splicing in uveal melanoma. Cancer Discov. 3, 1122–9.

Gale, R.E., Fielding, A.K., Harrison, C.N., Linch, D.C., 1997. Acquired skewing of X-chromosome inactivation patterns in myeloid cells of the elderly suggests stochastic clonal loss with age. Br. J. Haematol. 98, 512–519.

Gallipoli, P., Abraham, S.A., Holyoake, T.L., 2011. Hurdles Toward a Cure for CML: The CML Stem Cell. Hematol. Oncol. Clin. North Am. 25, 951–966.

Gangat, N., Caramazza, D., Vaidya, R., George, G., Begna, K., Schwager, S., Van Dyke, D., Hanson, C., Wu, W., Pardanani, A., et al., 2011. DIPSS Plus: A Refined Dynamic International Prognostic Scoring System for Primary Myelofibrosis That Incorporates Prognostic Information From Karyotype, Platelet Count, and Transfusion Status. J. Clin. Oncol. 29, 392–397.

Ge, H., Liu, K., Juan, T., Fang, F., Newman, M., Hoeck, W., 2011. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. Bioinformatics 27, 1922–1928.

Ge, Z., Guo, X., Li, J., Hartman, M., Kawasawa, Y.I., Dovat, S., Song, C., 2015. Clinical significance of high c-MYC and low MYCBP2 expression and their association with Ikaros dysfunction in adult acute lymphoblastic leukemia. Oncotarget 6, 42300–11.

Gisslinger, H., Zagrijtschuk, O., Buxhofer-Ausch, V., Thaler, J., Schloegl, E., Gastl, G.A., Wolf, D., Kralovics, R., Gisslinger, B., Strecker, K., et al., 2015. Ropeginterferon alfa-2b, a novel IFN -2b, induces high response rates with low toxicity in patients with polycythemia vera. Blood 126, 1762–1769.

Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al., 2012. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat. Genet. 44, 53–57.

Gregg, X.T., Kralovics, R., Prchal, J.T., 2000. A polymorphism of the X-linked gene IDS increases the number of females informative for transcriptional clonality assays. Am. J. Hematol. 63, 184–91.

Griesinger, F., Hennig, H., Hillmer, F., Podleschny, M., Steffens, R., Pies, A., Wörmann, B., Haase, D., Bohlander, S.K., 2005. ABCR-JAK2 fusion gene as the result of a t(9;22)(p24;q11.2) translocation in a patient with a clinically typical chronic myeloid leukemia. Genes, Chromosom. Cancer 44, 329–333.

Grinfeld, J., Nangalia, J., Baxter, E.J., Wedge, D.C., Angelopoulos, N., Cantrill, R., Godfrey, A.L., Papaemmanuil, E., Gundem, G., MacLean, C., et al., 2018. Classification and Personalized Prognosis in Myeloproliferative Neoplasms. N. Engl. J. Med. 379, 1416–1430.

Gross, S., Watson, L., Pathak, S., Khrebtukova, I., Schroth, G., 2014. Abstract 3562: Building workflows for gene fusion detection by RNA-seq. Cancer Res. 74, 3562-.

Gubin, M.M., Zhang, X., Schuster, H., Caron, E., Ward, J.P., Noguchi, T., Ivanova, Y., Hundal, J., Arthur, C.D., Krebber, W.-J., et al., 2014. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature 515, 577–81.

Guglielmelli, P., Rotunno, G., Bogani, C., Mannarelli, C., Giunti, L., Provenzano, A., Giglio, S., Squires, M., Stalbovskaya, V., Gopalakrishna, P., et al., 2016. Ruxolitinib is an effective treatment for *CALR* -positive patients with myelofibrosis. Br. J. Haematol. 173, 938–940.

Haider, M., Elala, Y.C., Gangat, N., Hanson, C.A., Tefferi, A., 2016. MPL mutations and palpable splenomegaly are independent risk factors for fibrotic progression in essential thrombocythemia. Blood Cancer J. 6, e487–e487.

Harbour, J.W., Roberson, E.D.O., Anbunathan, H., Onken, M.D., Worley, L.A., Bowcock, A.M., 2013. Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. Nat. Genet. 45, 133–135.

Harrison, C.N., Vannucchi, A.M., 2016. Closing the gap: genetic landscape of MPN. Blood 127, 276–278.

Harutyunyan, A.S., Kralovics, R., 2012. Role of Germline Genetic Factors in MPN Pathogenesis. Hematol. Oncol. Clin. North Am. 26, 1037–1051.

Hoeller, S., Walz, C., Reiter, A., Dirnhofer, S., Tzankov, A., 2011. PCM1–JAK2-fusion: a potential treatment target in myelodysplastic–myeloproliferative and other hemato-lymphoid neoplasms. Expert Opin. Ther. Targets 15, 53–62.

Holmström, M.O., Hjortsø, M.D., Ahmad, S.M., Met, Ö., Martinenaite, E., Riley, C., Straten, P., Svane, I.M., Hasselbalch, H.C., Andersen, M.H., 2017. The JAK2V617F mutation is a target for specific T cells in the JAK2V617F-positive myeloproliferative neoplasms. Leukemia 31, 495–498.

Holmström, M.O., Riley, C.H., Svane, I.M., Hasselbalch, H.C., Andersen, M.H., 2016. The CALR exon 9 mutations are shared neoantigens in patients with CALR mutant chronic myeloproliferative neoplasms. Leukemia 30, 2413–2416.

Holyoake, T., Jiang, X., Eaves, C., Eaves, A., 1999. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. Blood 94, 2056–64.

Hu, Z., Ott, P.A., Wu, C.J., 2017. Towards personalized, tumour-specific, therapeutic vaccines for cancer. Nat. Rev. Immunol. 18, 168–182.

Hussein, K., Van Dyke, D.L., Tefferi, A., 2009. Conventional cytogenetics in myelofibrosis: literature review and discussion. Eur. J. Haematol. 82, 329–338.

Ishii, T., Xu, M., Zhao, Y., Hu, W.-Y., Ciurea, S., Bruno, E., Hoffman, R., 2007. Recurrence of clonal hematopoiesis after discontinuing pegylated recombinant interferon-α 2a in a patient with polycythemia vera. Leukemia 21, 373–374.

Iyer, M.K., Chinnaiyan, A.M., Maher, C.A., 2011. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics 27, 2903–2904.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al., 2014. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. N. Engl. J. Med. 371, 2488–2498.

James, C., Ugo, V., Le Couédic, J.-P., Staerk, J., Delhommeau, F., Lacout, C., Garçon, L., Raslova, H., Berger, R., Bennaceur-Griscelli, A., et al., 2005. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature 434, 1144–1148.

Jia, W., Qiu, K., He, M., Song, P., Zhou, Q., Zhou, F., Yu, Y., Zhu, D., Nickerson, M.L., Wan, S., et al., 2013. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol. 14, R12.

Jin, Y., Mertens, F., Kullendorff, C.-M., Panagopoulos, Ioannis, 2006. Fusion of the Tumor-Suppressor Gene CHEK2 and the Gene for the Regulatory Subunit B of Protein Phosphatase 2 PPP2R2A in Childhood Teratoma. Neoplasia 8, 413–418.

Kennedy, J.A., Atenafu, E.G., Messner, H.A., Craddock, K.J., Brandwein, J.M., Lipton, J.H., Minden, M.D., Schimmer, A.D., Schuh, A.C., Yee, K.W., et al., 2013. Treatment outcomes following leukemic transformation in Philadelphia-negative myeloproliferative neoplasms. Blood 121, 2725–33.

Kiladjian, J.-J., Cassinat, B., Chevret, S., Turlure, P., Cambier, N., Roussel, M., Bellucci, S., Grandchamp, B., Chomienne, C., Fenaux, P., 2008. Pegylated interferon-alfa-2a induces complete hematologic and molecular responses with low toxicity in polycythemia vera. Blood 112, 3065–3072.

Kiladjian, J.-J., Chevret, S., Dosquet, C., Chomienne, C., Rain, J.-D., 2011. Treatment of Polycythemia Vera With Hydroxyurea and Pipobroman: Final Results of a Randomized Trial Initiated in 1980. J. Clin. Oncol. 29, 3907–3913.

Kiladjian, J.-J., Massé, A., Cassinat, B., Mokrani, H., Teyssandier, I., le Couédic, J.-P., Cambier, N., Almire, C., Pronier, E., Casadevall, N., et al., 2010. Clonal analysis of erythroid progenitors suggests that pegylated interferon α-2a treatment targets JAK2V617F clones without affecting TET2 mutant cells. Leukemia 24, 1519–1523.

Kim, D., Salzberg, S.L., 2011. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. 12, R72.

Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C.-W., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.-B., Murphy, M.E., et al., 2015a. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell 27, 617–630.

Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C.-W., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.-B., Murphy, M.E., et al., 2015b. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell 27, 617–630.

Kinsella, M., Harismendy, O., Nakano, M., Frazer, K.A., Bafna, V., 2011. Sensitive gene fusion detection using ambiguously mapping RNA-Seq read pairs. Bioinformatics 27, 1068–75.

Klampfl, T., Gisslinger, H., Harutyunyan, A.S., Nivarthi, H., Rumi, E., Milosevic, J.D., Them, N.C.C., Berg, T., Gisslinger, B., Pietra, D., et al., 2013. Somatic mutations of calreticulin in myeloproliferative neoplasms. N. Engl. J. Med. 369, 2379–90.

Klampfl, T., Harutyunyan, A., Berg, T., Gisslinger, B., Schalling, M., Bagienski, K., Olcaydu, D., Passamonti, F., Rumi, E., Pietra, D., et al., 2011. Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. Blood 118, 167–176.

Kralovics, R., 2008. Genetic complexity of myeloproliferative neoplasms. Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, U.K 22, 1841–1848.

Kralovics, R., Guan, Y., Prchal, J.T., 2002. Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. Exp. Hematol. 30, 229–36.

Kralovics, R., Passamonti, F., Buser, A.S., Teo, S., Tiedt, R., Passweg, J.R., Tichelli, A., Cazzola, M., Skoda, R.C., 2005. A Gain-of-Function Mutation of JAK2 in Myeloproliferative Disorders. N. Engl. J. Med. 352, 1779–1790.

Kreiter, S., Vormehr, M., van de Roemer, N., Diken, M., Löwer, M., Diekmann, J., Boegel, S., Schrörs, B., Vascotto, F., Castle, J.C., et al., 2015. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. Nature 520, 692–6.

Kumar, S., Vo, A.D., Qin, F., Li, H., 2016. Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. Sci. Rep. 6, 21597.

Kwong, Y.L., Chiu, E.K.W., Liang, R.H.S., Chan, V., Chan, T.K., 1996. Essential thrombocythemia with BCR/ABL rearrangement. Cancer Genet. Cytogenet. 89, 74–76.

Lacronique, V., Boureux, A., Valle, V.D., Poirel, H., Quang, C.T., Mauchauffé, M., Berthou, C., Lessard, M., Berger, R., Ghysdael, J., et al., 1997. A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. Science 278, 1309–12.

Lasho, T.L., Finke, C.M., Hanson, C.A., Jimma, T., Knudson, R.A., Ketterling, R.P.,

Pardanani, A., Tefferi, A., 2012. SF3B1 mutations in primary myelofibrosis: clinical, histopathology and genetic correlates among 155 patients. Leukemia 26, 1135–1137.

Lee, S.C.-W., North, K., Kim, E., Jang, E., Obeng, E., Lu, S.X., Liu, B., Inoue, D., Yoshimi, A., Ki, M., et al., 2018. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. Cancer Cell 34, 225–241.e8.

Lei, Q., Li, C., Zuo, Z., Huang, C., Cheng, H., Zhou, R., 2016. Evolutionary Insights into RNA trans-Splicing in Vertebrates. Genome Biol. Evol. 8, 562–77.

Lennerz, V., Fatho, M., Gentilini, C., Frye, R.A., Lifke, A., Ferel, D., Wolfel, C., Huber, C., Wolfel, T., 2005. The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. Proc. Natl. Acad. Sci. 102, 16013–16018.

Levine, R.L., Wadleigh, M., Cools, J., Ebert, B.L., Wernig, G., Huntly, B.J.P., Boggon, T.J., Wlodarska, I., Clark, J.J., Moore, S., et al., 2005. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell 7, 387–97.

Li, H., Wang, J., Mor, G., Sklar, J., 2008. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. Science 321, 1357–61.

Li, J., Kent, D.G., Chen, E., Green, A.R., 2011. Mouse models of myeloproliferative neoplasms: JAK of all grades. Dis. Model. Mech. 4, 311–7.

Li, X.-Q., Lu, J.-T., Tan, C.-C., Wang, Q.-S., Feng, Y.-M., 2016. RUNX2 promotes breast cancer bone metastasis by increasing integrin α5-mediated colonization. Cancer Lett. 380, 78–86.

Li, Y., Chien, J., Smith, D.I., Ma, J., 2011. FusionHunter: identifying fusion transcripts in cancer using paired-end RNA-seq. Bioinformatics 27, 1708–1710.

Lindberg, J., Ph, D., Rose, S. a, Bakhoum, S.F., Ph, D., Chambert, K., Mick, E., Neale, B.M., Ph, D., Fromer, M., et al., 2014. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. N. Engl. J. Med. 371, 2477–2487.

Linnemann, C., van Buuren, M.M., Bies, L., Verdegaal, E.M.E., Schotte, R., Calis, J.J.A., Behjati, S., Velds, A., Hilkmann, H., Atmioui, D. el, et al., 2015. High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. Nat. Med. 21, 81–85.

Liu, C., Ma, J., Chang, C., Zhou, X., 2013. FusionQ: a novel approach for gene fusion detection and quantification from paired-end RNA-Seq. BMC Bioinformatics 14, 193.

Liu, E., Jelinek, J., Pastore, Y.D., Guan, Y., Prchal, J.F.J.T.F., Prchal, J.F.J.T.F., 2003. Discrimination of polycythemias and thrombocytoses by novel, simple, accurate clonality assays and comparison with PRV-1 expression and BFU-E response to erythropoietin. Blood 101, 3294–3301.

Liu, S., Tsai, W.-H., Ding, Y., Chen, R., Fang, Z., Huo, Z., Kim, S., Ma, T., Chang, T.-Y., Priedigkeit, N.M., et al., 2016. Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. Nucleic Acids Res. 44, e47.

Lu, X., Levine, R., Tong, W., Wernig, G., Pikman, Y., Zarnegar, S., Gilliland, D.G., Lodish, H., 2005. Expression of a homodimeric type I cytokine receptor is required for JAK2V617F-mediated transformation. Proc. Natl. Acad. Sci. 102, 18962–18967.

Lundberg, P., Karow, A., Nienhold, R., Looser, R., Hao-Shen, H., Nissen, I., Girsberger, S., Lehmann, T., Passweg, J., Stern, M., et al., 2014a. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood 123, 2220–8.

Lundberg, P., Takizawa, H., Kubovcakova, L., Guo, G., Hao-Shen, H., Dirnhofer, S., Orkin, S.H., Manz, M.G., Skoda, R.C., 2014b. Myeloproliferative neoplasms can be initiated from a single hematopoietic stem cell expressing JAK2-V617F. J. Exp. Med. 211, 2213–30.

Lyon, M.F., 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190, 372–3.

Maguire, S.L., Leonidou, A., Wai, P., Marchiò, C., Ng, C.K., Sapino, A., Salomon, A.-V., Reis-Filho, J.S., Weigelt, B., Natrajan, R.C., 2015. *SF3B1* mutations constitute a novel therapeutic target in breast cancer. J. Pathol. 235, 571–580.

Marchioli, R., Finazzi, G., Landolfi, R., Kutti, J., Gisslinger, H., Patrono, C., Marilus, R., Villegas, A., Tognoni, G., Barbui, T., 2005. Vascular and Neoplastic Risk in a Large Cohort of Patients With Polycythemia Vera. J. Clin. Oncol. 23, 2224–2232.

Marty, C., Pecquet, C., Nivarthi, H., El-Khoury, M., Chachoua, I., Tulliez, M., Villeval, J.-L., Raslova, H., Kralovics, R., Constantinescu, S.N., et al., 2016. Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. Blood 127, 1317–1324.

Massaro, P., Foa, P., Pomati, M., LaTargia, M.L., Iurlo, A., Clerici, C., Caldiera, S., Fornier, M., Maiolo, A.T., 1997. Polycythemia vera treated with recombinant interferon-alpha 2a: evidence of a selective effect on the malignant clone. Am. J. Hematol. 56, 126–8.

Matsushita, H., Vesely, M.D., Koboldt, D.C., Rickert, C.G., Uppaluri, R., Magrini, V.J., Arthur, C.D., White, J.M., Chen, Y.-S., Shea, L.K., et al., 2012. Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. Nature 482, 400–404.

McNerney, M.E., Brown, C.D., Wang, X., Bartom, E.T., Karmakar, S., Bandlamudi, C., Yu, S., Ko, J., Sandall, B.P., Stricker, T., et al., 2013. CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia. Blood 121, 975–983.

McPherson, A., Hormozdiari, F., Zayed, A., Giuliany, R., Ha, G., Sun, M.G.F., Griffith, M., Moussavi, A., Senz, J., Melnyk, N., et al., 2011a. Defuse: An algorithm for gene fusion discovery in tumor rna-seq data. PLoS Comput. Biol. 7.

McPherson, A., Wu, C., Hajirasouliha, I., Hormozdiari, F., Hach, F., Lapuk, A., Volik, S., Shah, S., Collins, C., Sahinalp, S.C., 2011b. Comrad: detection of expressed rearrangements by integrated analysis of RNA-Seq and low coverage genome sequence data. Bioinformatics 27, 1481–1488.

Melo, S.A., Moutinho, C., Ropero, S., Calin, G.A., Rossi, S., Spizzo, R., Fernandez, A.F., Davalos, V., Villanueva, A., Montoya, G., et al., 2010. A Genetic Defect in Exportin-5 Traps Precursor MicroRNAs in the Nucleus of Cancer Cells. Cancer Cell 18, 303–

315.

Mertens, F., Johansson, B., Fioretos, T., Mitelman, F., 2015. The emerging complexity of gene fusions in cancer. Nat. Rev. Cancer 15, 371–381.

Mian, S.A., Rouault-Pierre, K., Smith, A.E., Seidl, T., Pizzitola, I., Kizilors, A., Kulasekararaj, A.G., Bonnet, D., Mufti, G.J., 2015. SF3B1 mutant MDS-initiating cells may arise from the haematopoietic stem cell compartment. Nat. Commun. 6, 10004.

Milosevic Feenstra, J.D., Nivarthi, H., Gisslinger, H., Leroy, E., Rumi, E., Chachoua, I., Bagienski, K., Kubesova, B., Pietra, D., Gisslinger, B., et al., 2016. Whole-exome sequencing identifies novel MPL and JAK2 mutations in triple-negative myeloproliferative neoplasms. Blood 127, 325–332.

Milosevic, J.D., Puda, A., Malcovati, L., Berg, T., Hofbauer, M., Stukalov, A., Klampfl, T., Harutyunyan, A.S., Gisslinger, H., Gisslinger, B., et al., 2012. Clinical significance of genetic aberrations in secondary acute myeloid leukemia. Am. J. Hematol. 87, 1010–6.

Mitelman, F., Johansson, B., Mertens, F., 2013. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer.

Mitelman, F., Johansson, B., Mertens, F., 2007. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 7, 233–245.

Murati, A., Adélaïde, J., Gelsi-Boyer, V., Etienne, A., Rémy, V., Fezoui, H., Sainty, D., Xerri, L., Vey, N., Olschwang, S., et al., 2006. t(5;12)(q23–31;p13) with ETV6-ACSL6 gene fusion in polycythemia vera. Leukemia 20, 1175–1178.

Nangalia, J., Massie, C.E., Baxter, E.J., Nice, F.L., Gundem, G., Wedge, D.C., Avezov, E., Li, J., Kollmann, K., Kent, D.G., et al., 2013. Somatic *CALR* Mutations in Myeloproliferative Neoplasms with Nonmutated *JAK2*. N. Engl. J. Med. 369, 2391–2405.

Nicorici, D., Satalan, M., Edgren, H., Kangaspeska, S., Murumagi, A., Kallioniemi, O., Virtanen, S., Kilkku, O., 2014. FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data, bioRxiv. Cold Spring Harbor Labs Journals.

Nivarthi, H., Chen, D., Cleary, C., Kubesova, B., Jäger, R., Bogner, E., Marty, C., Pecquet, C., Vainchenker, W., Constantinescu, S.N., et al., 2016. Thrombopoietin receptor is required for the oncogenic function of CALR mutants. Leukemia 30, 1759–1763.

Novellino, L., Castelli, C., Parmiani, G., 2005. A listing of human tumor antigens recognized by T cells: March 2004 update. Cancer Immunol. Immunother. 54, 187–207.

Obeng, E.A., Chappell, R.J., Seiler, M., Chen, M.C., Campagna, D.R., Schmidt, P.J., Schneider, R.K., Lord, A.M., Wang, L., Gambe, R.G., et al., 2016. Physiologic Expression of Sf3b1K700E Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. Cancer Cell 30, 404–417.

Onodera, Y., Miki, Y., Suzuki, T., Takagi, K., Akahira, J., Sakyu, T., Watanabe, M., Inoue, S., Ishida, T., Ohuchi, N., et al., 2010. Runx2 in human breast carcinoma: its potential roles in cancer progression. Cancer Sci. 101, 2670–2675.

Ortmann, C. a., Kent, D.G., Nangalia, J., Silber, Y., Wedge, D.C., Grinfeld, J., Baxter, E.J., Massie, C.E., Papaemmanuil, E., Menon, S., et al., 2015. Effect of Mutation Order on Myeloproliferative Neoplasms. N. Engl. J. Med. 372, 601–612.

Ott, P.A., Hu, Z., Keskin, D.B., Shukla, S.A., Sun, J., Bozym, D.J., Zhang, W., Luoma, A., Giobbie-Hurder, A., Peter, L., et al., 2017. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature 547, 217–221.

Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., et al., 2011. Somatic *SF3B1* Mutation in Myelodysplasia with Ring Sideroblasts. N. Engl. J. Med. 365, 1384–1395.

Pardanani, A.D., Levine, R.L., Lasho, T., Pikman, Y., Mesa, R.A., Wadleigh, M., Steensma, D.P., Elliott, M.A., Wolanskyj, A.P., Hogan, W.J., et al., 2006. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. Blood 108, 3472–3476.

Peeters, P., Raynaud, S.D., Cools, J., Wlodarska, I., Grosgeorge, J., Philip, P., Monpoux, F., Van Rompaey, L., Baens, M., Van den Berghe, H., et al., 1997. Fusion of TEL, the ETS-variant gene 6 (ETV6), to the receptor-associated kinase JAK2 as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia. Blood 90, 2535–40.

Pellagatti, A., Armstrong, R.N., Steeples, V., Sharma, E., Repapi, E., Singh, S., Sanchi, A., Radujkovic, A., Horn, P., Dolatshad, H., et al., 2018. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. Blood blood-2018-04-843771.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., Brockdorff, N., 1996. Requirement for Xist in X chromosome inactivation. Nature 379, 131–137.

Pereira, C., Gimenez-Xavier, P., Pros, E., Pajares, M.J., Moro, M., Gomez, A., Navarro, A., Condom, E., Moran, S., Gomez-Lopez, G., et al., 2017. Genomic Profiling of Patient-Derived Xenografts for Lung Cancer Identifies B2M Inactivation Impairing Immunorecognition. Clin. Cancer Res. 23, 3203–3213.

Piazza, R., Pirola, A., Spinelli, R., Valletta, S., Redaelli, S., Magistroni, V., Gambacorti-Passerini, C., 2012. FusionAnalyser: a new graphical, event-driven tool for fusion rearrangements discovery. Nucleic Acids Res. 40, e123.

Pikman, Y., Lee, B.H., Mercher, T., McDowell, E., Ebert, B.L., Gozo, M., Cuker, A., Wernig, G., Moore, S., Galinsky, I., et al., 2006. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med. 3, e270.

Plo, I., Zhang, Y., Le Couédic, J.-P., Nakatake, M., Boulet, J.-M., Itaya, M., Smith, S.O., Debili, N., Constantinescu, S.N., Vainchenker, W., et al., 2009. An activating mutation in the *CSF3R* gene induces a hereditary chronic neutrophilia. J. Exp. Med. 206, 1701–1707.

Prchal, J.T., Guan, Y.L., 1993. A novel clonality assay based on transcriptional analysis of the active X chromosome. Stem Cells 11, 62–65.

Qiu, M.-T., Fan, Q., Zhu, Z., Kwan, S.-Y., Chen, L., Chen, J.-H., Ying, Z.-L., Zhou, Y., Gu, W., Wang, L.-H., et al., 2015. KDM4B and KDM4A promote endometrial cancer progression by regulating androgen receptor, c-myc, and p27<sup>kip1</sup> Oncotarget 6, 31702–20.

Quesada, V., Conde, L., Villamor, N., Ordóñez, G.R., Jares, P., Bassaganyas, L., Ramsay, A.J., Beà, S., Pinyol, M., Martínez-Trillos, A., et al., 2012. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat. Genet. 44, 47–52.

Quintás-Cardama, A., Kantarjian, H., Manshouri, T., Luthra, R., Estrov, Z., Pierce, S., Richie, M.A., Borthakur, G., Konopleva, M., Cortes, J., et al., 2009. Pegylated interferon alfa-2a yields high rates of hematologic and molecular response in patients with advanced essential thrombocythemia and polycythemia vera. J. Clin. Oncol. 27, 5418–24.

Quivoron, C., Couronné, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.-H., et al., 2011. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. Cancer Cell 20, 25–38.

Rajasagi, M., Shukla, S.A., Fritsch, E.F., Keskin, D.B., DeLuca, D., Carmona, E., Zhang, W., Sougnez, C., Cibulskis, K., Sidney, J., et al., 2014. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. Blood 124, 453–462.

Rampal, R., Al-Shahrour, F., Abdel-Wahab, O., Patel, J.P., Brunel, J.-P., Mermel, C.H., Bass, A.J., Pretz, J., Ahn, J., Hricik, T., et al., 2014. Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. Blood 123, e123-33.

Rice, L., Popat, U., 2005. Every case of essential thrombocythemia should be tested for the Philadelphia chromosome. Am. J. Hematol. 78, 71–73.

Robbins, P.F., Lu, Y.-C., El-Gamil, M., Li, Y.F., Gross, C., Gartner, J., Lin, J.C., Teer, J.K., Cliften, P., Tycksen, E., et al., 2013. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. Nat. Med. 19, 747–52.

Robert, L., Tsoi, J., Wang, X., Emerson, R., Homet, B., Chodon, T., Mok, S., Huang, R.R., Cochran, A.J., Comin-Anduix, B., et al., 2014. CTLA4 Blockade Broadens the Peripheral T-Cell Receptor Repertoire. Clin. Cancer Res. 20, 2424–2432.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., et al., 2010. De novo assembly and analysis of RNA-seq data. Nat. Methods 7, 909–912.

Rodriguez-Perales, S., Torres-Ruiz, R., Suela, J., Acquadro, F., Martin, M.C., Yebra, E., Ramirez, J.C., Alvarez, S., Cigudosa, J.C., 2015. Truncated RUNX1 protein generated by a novel t(1;21)(p32;q22) chromosomal translocation impairs the proliferation and differentiation of human hematopoietic progenitors. Oncogene.

Rossi, D., Bruscaggin, A., Spina, V., Rasi, S., Khiabanian, H., Messina, M., Fangazio, M., Vaisitti, T., Monti, S., Chiaretti, S., et al., 2011. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood 118, 6904–6908.

Rotunno, G., Mannarelli, C., Guglielmelli, P., Pacilli, A., Pancrazzi, A., Pieri, L., Fanelli, T., Bosi, A., Vannucchi, A.M., Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative Investigators, 2014. Impact of calreticulin mutations on clinical and hematological phenotype and outcome in

essential thrombocythemia. Blood 123, 1552–1555.

Rumi, E., Cazzola, M., 2017. Advances in understanding the pathogenesis of familial myeloproliferative neoplasms. Br. J. Haematol. 178, 689–698.

Rumi, E., Pietra, D., Ferretti, V., Klampfl, T., Harutyunyan, A., Milosevic, J., Them, N., Berg, T., Elena, C., Casetti, I., et al., 2014a. JAK2 or CALR mutation status defines subtypes of essential thrombocythemia with substantially different clinical course and outcomes. Blood 123, 1544–1551.

Rumi, E., Pietra, D., Guglielmelli, P., Bordoni, R., Casetti, I., Milanesi, C., Sant'Antonio, E., Ferretti, V., Pancrazzi, A., Rotunno, G., et al., 2013. Acquired copy-neutral loss of heterozygosity of chromosome 1p as a molecular event associated with marrow fibrosis in MPL-mutated myeloproliferative neoplasms. Blood 121, 4388–95.

Rumi, E., Pietra, D., Pascutto, C., Guglielmelli, P., Martínez-Trillos, A., Casetti, I., Colomer, D., Pieri, L., Pratcorona, M., Rotunno, G., et al., 2014b. Clinical effect of driver mutations of JAK2, CALR, or MPL in primary myelofibrosis. Blood 124, 1062–9.

Rumi, E., Sant'Antonio, E., Boveri, E., Pietra, D., Cavalloni, C., Roncoroni, E., Astori, C., Arcaini, L., 2018. Diagnosis and management of prefibrotic myelofibrosis. Expert Rev. Hematol. 11, 537–545.

Saeterdal, I., Bjorheim, J., Lislerud, K., Gjertsen, M.K., Bukholm, I.K., Olsen, O.C., Nesland, J.M., Eriksen, J.A., Moller, M., Lindblom, A., et al., 2001. Frameshift-mutation-derived peptides as tumor-specific antigens in inherited and spontaneous colorectal cancer. Proc. Natl. Acad. Sci. 98, 13255–13260.

Saharinen, P., Silvennoinen, O., 2002. The pseudokinase domain is required for suppression of basal activity of Jak2 and Jak3 tyrosine kinases and for cytokine-inducible activation of signal transduction. J. Biol. Chem. 277, 47954–63.

Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.-P., Simon, P., Löwer, M., Bukur, V., Tadmor, A.D., Luxemburger, U., Schrörs, B., et al., 2017. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature 547, 222–226.

Sarris, M.E., Moulos, P., Haroniti, A., Giakountis, A., Talianidis, I., 2016. Smyd3 Is a Transcriptional Potentiator of Multiple Cancer-Promoting Genes and Required for Liver and Colon Cancer Development. Cancer Cell 29, 354–366.

Sboner, A., Habegger, L., Pflueger, D., Terry, S., Chen, D.Z., Rozowsky, J.S., Tewari, A.K., Kitabayashi, N., Moss, B.J., Chee, M.S., et al., 2010. FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol. 11, R104.

Schaub, F.X., Looser, R., Li, S., Hao-Shen, H., Lehmann, T., Tichelli, A., Skoda, R.C., 2010. Clonal analysis of TET2 and JAK2 mutations suggests that TET2 can be a late event in the progression of myeloproliferative neoplasms. Blood 115, 2003–2007.

Schischlik, F., Kralovics, R., 2017. Mutations in myeloproliferative neoplasms – their significance and clinical use. Expert Rev. Hematol. 10, 961–973.

Schumacher, T.N., Schreiber, R.D., 2015. Neoantigens in cancer immunotherapy. Science 348, 69–74.

Scott, L.M., 2011. The JAK2 exon 12 mutations: A comprehensive review. Am. J. Hematol. 86, 668–676.

Scott, L.M., Tong, W., Levine, R.L., Scott, M.A., Beer, P.A., Stratton, M.R., Futreal, P.A., Erber, W.N., McMullin, M.F., Harrison, C.N., et al., 2007. *JAK2* Exon 12 Mutations in Polycythemia Vera and Idiopathic Erythrocytosis. N. Engl. J. Med. 356, 459–468.

Seghatoleslami, M., Ketabchi, N., Ordo, A., Asl, J.M., Golchin, N., Saki, N., 2016. Coexistence of p190 BCR/ABL Transcript and CALR 52-bp Deletion in Chronic Myeloid Leukemia Blast Crisis: A Case Report. Mediterr. J. Hematol. Infect. Dis. 8, e2016002.

Shigehisa Kitano, A.I., Kitano, S., Kim, Y., Inoue, M., Fuse, M., Tada, K., Yoshimura, K., 2015. Cancer Neoantigens: A Promising Source of Immunogens for Cancer Immunotherapy. J. Clin. Cell. Immunol. 06, 1–7.

Shigeyasu, K., Okugawa, Y., Toden, S., Boland, C.R., Goel, A., 2017. Exportin-5 Functions as an Oncogene and a Potential Therapeutic Target in Colorectal Cancer. Clin. Cancer Res. 23, 1312–1322.

Shreenivas, A., Mascarenhas, J., 2018. Emerging drugs for the treatment of Myelofibrosis. Expert Opin. Emerg. Drugs 23, 37–49.

Slavney, A., Arbiza, L., Clark, A.G., Keinan, A., 2015. Strong Constraint on Human Genes Escaping X-Inactivation Is Modulated by their Expression Level and Breadth in Both Sexes. Mol. Biol. Evol. msv225-.

Snyder, A., Chan, T.A., 2015. Immunogenic peptide discovery in cancer genomes. Curr. Opin. Genet. Dev. 30, 7–16.

Soret, J., Cassinat, B., Chevret, S., Dosquet, C., Giraudier, S., Verger, E., Maslah, N., Parquet, N., Chomienne, C., Kiladjian, J.-J., 2016. Outcomes of Patients with Myeloproliferative Neoplasms (MPN) after Interferon-Alpha (IFN) Therapy Discontinuation. Blood 128.

Staerk, J., Lacout, C., Sato, T., Smith, S.O., Vainchenker, W., Constantinescu, S.N., 2006. An amphipathic motif at the transmembrane-cytoplasmic junction prevents autonomous activation of the thrombopoietin receptor. Blood 107, 1864–1871.

Stegelmann, F., Bullinger, L., Schlenk, R.F., Paschka, P., Griesshammer, M., Blersch, C., Kuhn, S., Schauer, S., Döhner, H., Döhner, K., 2011. DNMT3A mutations in myeloproliferative neoplasms. Leukemia 25, 1217–1219.

Storlazzi, C.T., Von Steyern, F.V., Domanski, H.A., Mandahl, N., Mertens, F., 2005. Biallelic somatic inactivation of theNF1 gene through chromosomal translocations in a sporadic neurofibroma. Int. J. Cancer 117, 1055–1057.

Swerdlow, S.H., Campo, E., Harris Nancy Lee, Jaffe Elaine S, Pileri Stefano A, Stein Harald, Thiele Jürgen, 2017. WHO classification of tumours of haematopoietic and lymphoid tissues (Revised 4th edition), 4th ed. Lyon.

Swierczek, S.I., Piterkova, L., Jelinek, J., Agarwal, N., Hammoud, S., Wilson, A., Hickman, K., Parker, C.J., Cairns, B.R., Prchal, J.T., et al., 2012. Methylation of AR locus does not always reflect X chromosome inactivation state. Blood 119, e100–e109.

Szelinger, S., Malenica, I., Corneveaux, J.J., Siniard, A.L., Kurdoglu, A.A., Ramsey, K.M., Schrauwen, I., Trent, J.M., Narayanan, V., Huentelman, M.J., et al., 2014. Characterization of X chromosome inactivation using integrated analysis of whole-exome and mRNA sequencing. PLoS One 9, e113036.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–5.

Tanaka, H., Takeuchi, M., Takeda, Y., Sakai, S., Abe, D., Ohwada, C., Sakaida, E., Shimizu, N., Saito, Y., Miyagi, S., et al., 2010. Identification of a novel TEL–Lyn fusion gene in primary myelofibrosis. Leukemia 24, 197–200.

Taylor, B.S., Ladanyi, M., 2011. Clinical cancer genomics: how soon is now? J. Pathol. 223, 319–327.

Tefferi, A., 2010. Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms: JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1. Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, U.K 24, 1128–1138.

Tefferi, A., Finke, C.M., Lasho, T.L., Hanson, C.A., Ketterling, R.P., Gangat, N., Pardanani, A., 2018a. U2AF1 mutation types in primary myelofibrosis: phenotypic and prognostic distinctions. Leukemia 32, 2274–2278.

Tefferi, A., Guglielmelli, P., Pardanani, A., Vannucchi, A.M., 2018b. Myelofibrosis Treatment Algorithm 2018. Blood Cancer J. 8, 72.

Tefferi, A., Lasho, T.L., Finke, C., Belachew, A.A., Wassie, E.A., Ketterling, R.P., Hanson, C.A., Pardanani, A., 2014. Type 1 vs type 2 calreticulin mutations in primary myelofibrosis: differences in phenotype and prognostic impact. Leukemia 28, 1568–1570.

Tefferi, A., Lasho, T.L., Finke, C.M., Knudson, R.A., Ketterling, R., Hanson, C.H., Maffioli, M., Caramazza, D., Passamonti, F., Pardanani, A., 2014a. CALR vs JAK2 vs MPL-mutated or triple-negative myelofibrosis: clinical, cytogenetic and molecular comparisons. Leukemia 28, 1472–1477.

Tefferi, A., Lasho, T.L., Tischer, A., Wassie, E.A., Finke, C.M., Belachew, A.A., Ketterling, R.P., Hanson, C.A., Pardanani, A.D., 2014b. The prognostic advantage of calreticulin mutations in myelofibrosis might be confined to type 1 or type 1-like CALR variants. Blood 124, 2465–6.

Tefferi, A., Rumi, E., Finazzi, G., Gisslinger, H., Vannucchi, A.M., Rodeghiero, F., Randi, M.L., Vaidya, R., Cazzola, M., Rambaldi, A., et al., 2013. Survival and prognosis among 1545 patients with contemporary polycythemia vera: an international study. Leukemia 27, 1874–1881.

Tefferi, A., Vannucchi, A.M., Barbui, T., 2018c. Essential thrombocythemia treatment algorithm 2018. Blood Cancer J. 8, 2.

Tefferi, A., Vannucchi, A.M., Barbui, T., 2018d. Polycythemia vera treatment algorithm 2018. Blood Cancer J. 8, 3.

Tefferi, A., Wassie, E.A., Guglielmelli, P., Gangat, N., Belachew, A.A., Lasho, T.L., Finke, C., Ketterling, R.P., Hanson, C.A., Pardanani, A., et al., 2014c. Type 1 versus Type 2 calreticulin mutations in essential thrombocythemia: a collaborative study of

1027 patients. Am. J. Hematol. 89, E121-4.

Them, N.C.C., Bagienski, K., Berg, T., Gisslinger, B., Schalling, M., Chen, D., Buxhofer-Ausch, V., Thaler, J., Schloegl, E., Gastl, G.A., et al., 2015. Molecular responses and chromosomal aberrations in patients with polycythemia vera treated with peg-proline-interferon alpha-2b. Am. J. Hematol. 90, 288–94.

Theobald, M., Biggs, J., Hernández, J., Lustgarten, J., Labadie, C., Sherman, L.A., 1997. Tolerance to p53 by A2.1-restricted cytotoxic T lymphocytes. J. Exp. Med. 185, 833–41.

Tiedt, R., Hao-Shen, H., Sobas, M.A., Looser, R., Dirnhofer, S., Schwaller, J., Skoda, R.C., 2008. Ratio of mutant JAK2-V617F to wild-type Jak2 determines the MPD phenotypes in transgenic mice. Blood 111, 3931–3940.

Titmarsh, G.J., Duncombe, A.S., McMullin, M.F., O'Rorke, M., Mesa, R., De Vocht, F., Horan, S., Fritschi, L., Clarke, M., Anderson, L.A., 2014. How common are myeloproliferative neoplasms? A systematic review and meta-analysis. Am. J. Hematol. 89, 581–7.

Tiziana Storlazzi, C., Pieri, L., Paoli, C., Daniele, G., Lasho, T., Tefferi, A., Vannucchi, A.M., 2014. Complex karyotype in a polycythemia vera patient with a novel SETD1B/GTF2H3 fusion gene. Am. J. Hematol. 89, 438–42.

Tran, E., Turcotte, S., Gros, A., Robbins, P.F., Lu, Y.-C., Dudley, M.E., Wunderlich, J.R., Somerville, R.P., Hogan, K., Hinrichs, C.S., et al., 2014. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. Science 344, 641–5.

Türeci, O., Vormehr, M., Diken, M., Kreiter, S., Huber, C., Sahin, U., 2016. Targeting the Heterogeneity of Cancer with Individualized Neoepitope Vaccines. Clin. Cancer Res. 22, 1885–1896.

Turlure, P., Cambier, N., Roussel, M., Bellucci, S., Zini, J.-M., Rain, J.-D., Rousselot, P., Vainchenker, W., Abdelkader, L., Ghomari, K., et al., 2015. Complete Hematological, Molecular and Histological Remissions without Cytoreductive Treatment Lasting After Pegylated-Interferon α-2a (peg-IFNα-2a) Therapy in Polycythemia Vera (PV): Long Term Results of a Phase 2 Trial. Blood 118, 280.

Vainchenker, W., Kralovics, R., 2017. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. Blood 129, 667–679.

van Rooij, N., van Buuren, M.M., Philips, D., Velds, A., Toebes, M., Heemskerk, B., van Dijk, L.J.A., Behjati, S., Hilkmann, H., El Atmioui, D., et al., 2013. Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. J. Clin. Oncol. 31, e439-42.

Vannucchi, A.M., Harrison, C.N., 2017. Emerging treatments for classical myeloproliferative neoplasms. Blood 129, 693–703.

Vannucchi, A.M., Lasho, T.L., Guglielmelli, P., Biamonte, F., Pardanani, A., Pereira, A., Finke, C., Score, J., Gangat, N., Mannarelli, C., et al., 2013. Mutations and prognosis in primary myelofibrosis. Leukemia 27, 1861–9.

Verger, E., Cassinat, B., Chauveau, A., Dosquet, C., Giraudier, S., Schlageter, M.-H., Ianotto, J.-C., Yassin, M.A., Al-Dewik, N., Carillo, S., et al., 2015. Clinical and molecular response to interferon-α therapy in essential thrombocythemia patients

with CALR mutations. Blood 126, 2585–91.

Wang, K., Li, M., Hakonarson, H., 2010a. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S. a., Perou, C.M., et al., 2010b. MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 38, 1–14.

Wang, L., Lawrence, M.S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D.S., Zhang, L., et al., 2011. *SF3B1* and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. N. Engl. J. Med. 365, 2497–2506.

Wang, Q., Xia, J., Jia, P., Pao, W., Zhao, Z., 2013. Application of next generation sequencing to human gene fusion detection: Computational tools, features and perspectives. Brief. Bioinform. 14, 506–519.

Welch, J.S., Westervelt, P., Ding, L., Larson, D.E., Klco, J.M., Kulkarni, S., Wallis, J., Chen, K., Payton, J.E., Fulton, R.S., et al., 2011. Use of Whole-Genome Sequencing to Diagnose a Cryptic Fusion Oncogene. JAMA 305, 1577.

Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., et al., 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. 20, 1472–8.

Yamada, O., Mahfoudhi, E., Plo, I., Ozaki, K., Nakatake, M., Akiyama, M., Yamada, H., Kawauchi, K., Vainchenker, W., 2014. Emergence of a BCR-ABL translocation in a patient with the JAK2V617F mutation: evidence for secondary acquisition of BCR-ABL in the JAK2V617F clone. J. Clin. Oncol. 32, e76-9.

Yizhak, K., Aguet, F., Kim, J., Hess, J., Kubler, K., Grimsby, J., Frazer, R., Zhang, H., Haradhvala, N., Rosebrock, D., et al., 2018. A comprehensive analysis of RNA sequences reveals macroscopic somatic clonal expansion across normal tissues. bioRxiv 416339.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al., 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478, 64–9.

Yoshihara, K., Wang, Q., Torres-Garcia, W., Zheng, S., Vegesna, R., Kim, H., Verhaak, R.G.W., 2014. The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene.

Zhang, G.L., Ansari, H.R., Bradley, P., Cawley, G.C., Hertz, T., Hu, X., Jojic, N., Kim, Y., Kohlbacher, O., Lund, O., et al., 2011. Machine learning competition in immunology – Prediction of HLA class I binding peptides. J. Immunol. Methods 374, 1–4.

Zhang, H., Chen, J., 2018. Current status and future directions of cancer immunotherapy. J. Cancer 9, 1773–1781.

Zhang, S.-J., Rampal, R., Manshouri, T., Patel, J., Mensah, N., Kayserian, A., Hricik, T., Heguy, A., Hedvat, C., Gonen, M., et al., 2012. Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. Blood 119, 4480–4485.

Zhang, X., Su, J., Jeong, M., Ko, M., Huang, Y., Park, H.J., Guzman, A., Lei, Y., Huang, Y.-H., Rao, A., et al., 2016. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. Nat. Genet. 48, 1014–1023.

Zheng, S., Torres-garcia, W., Biology, C., 2012. Identification of Gene Fusions Using RNA Sequencing Data PRADA.

Zhou, A., Knoche, E.M., Engle, E.K., Fisher, D.A.C., Oh, S.T., 2015. Concomitant JAK2 V617F-positive polycythemia vera and BCR-ABL-positive chronic myelogenous leukemia treated with ruxolitinib and dasatinib. Blood Cancer J. 5, e351.

7 ABBREVIATIONS

ABL1	Abelson murine leukemia viral oncogene homolog 1
ACOT11	Acyl-CoA Thioesterase 11
ACT	Adoptive T cell therapy
AHCTF1	AT-Hook Containing Transcription Factor 1
AML	Acute myeloid leukemia
AR	Androgen receptor
ASCT	Allogeneic stem cell transplantation
ASXL1	Additional sex comb like 1
B2M	Beta-2-Microglobulin
Ba/F3	A murine pro-B cell line
BCR	Breakpoint cluster region
BRCA	Breast cancer
C7orf43	Chromosome 7 Open Reading Frame 43
CALR	Calreticulin
CAR	Chimeric antigen receptor
CBFB	Core-Binding Factor Subunit Beta
CBL	Cbl Proto-Oncogene
CBY1	Chibby Family Member 1, Beta Catenin Antagonist
CC2D1A	Coiled-Coil And C2 Domain Containing 1A
CD	Cluster of differentiation
CDR	Common deleted region
CEL	Chronic esosinophilic leukemia
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
CMML	Chronic myelomonocytic leukemia
CNL	Chronic neutrophilic leukemia
CPB	Checkpoint blockade therapy
CSF3R	Colony Stimulating Factor 3 Receptor
CTL	Cytotoxic T cells
CTSD	Cathepsin D
CUX1	Cut-like homeobox 1
CYHR1	Cysteine and Histidine Rich 1
dbSNP	Single Nucleotide Polymorphism Database
DIPSS	Dynamic international prognostic scoring system
DNMT3A	DNA (cytosine-5)-methyltransferase 3 alpha
DP	Depth
EFEMP2	EGF Containing Fibulin Extracellular Matrix Protein 2
EPO	Erythropoietin
ET	Essential thrombocythemia

EZH2	Enhancer of zeste homolog 2
FDA	Food and Drug Administration
FGFR1	Fibroblast Growth Factor Receptor 1
FISH	Fluorescence In Situ Hybridization
FLT3LG	Fms Related Tyrosine Kinase 3 Ligand
FOXO3	Forkhead Box O3
FOXP1	Forkhead Box P1
FRY	FRY Microtubule Binding Protein
G6PD	Glucose-6-phosphate dehydrogenase (G6PD)
GATK	Genome Analysis Toolkit
GPR1	G Protein-Coupled Receptor 1
HDAC	Histone Deacetylase
Hi-C	Chromosome-wide conformation capture
HINT2	Histidine Triad Nucleotide Binding Protein 2
HLA	Human Leukocyte Antigen
HSC	Hematopoietic stem cell
HT	Hereditary thrombocytosis
HUMARA	Human androgen receptor gene
IDH1	Isocitrate dehydrogenase 1
IDH2	Isocitrate dehydrogenase 2
Indel	Insertion and deletion
INF-α	Interferon-α
INO80D	INO80 Complex Subunit D
inv	Inversion
JAK1	Janus kinase 1
JAK2	Janus kinase 2
JAK3	Janus kinase 3
KDM4B	Lysine Demethylase 4B
LOLX2	Lysyl oxidase like 2
mAbs	monoclonal Antibodies
MAP3K7	Mitogen-Activated Protein Kinase Kinase Kinase 7
MDS	Myelodysplastic dyndrome
MF	Myelofibrosis
MHC	Major Histocompatibility Complex
MLL	Lysine Methyltransferase 2A
MNDA	Myeloid Cell Nuclear Differentiation Antigen
MPM	Malignant Pleural Mesothelioma
MPN	Myeloproliferative neoplasms
MPNu	unclassifiable MPN
MYCBP2	MYC Binding Protein 2, E3 Ubiquitin Protein Ligase
MYH11	Myosin Heavy Chain 11
NMD	Nonsense mediated decay
NOX5	NADPH Oxidase 5
OXA1L	OXA1L, Mitochondrial Inner Membrane Protein
PAR1 & 2	Pseudo-autosomal regions 1 & 2

PC	Principal component
PDGFRA	Platelet Derived Growth Factor Receptor Alpha
PDGFRB	Platelet Derived Growth Factor Receptor Beta
PET-PV	Post-ET PV
PICALM	Phosphatidylinositol Binding Clathrin Assembly Protein
PLEK	Pleckstrin
PMF	Primary myelofibrosis
PML	Promyelocytic Leukemia
PNISR	PNN Interacting Serine And Arginine Rich Protein
PPM1M	Protein Phosphatase, Mg2+/Mn2+ Dependent 1M
PPV	Post-PV
PRPF38A	Pre-MRNA Processing Factor 38A
PV	Polycythemia vera
QD	Quality by depth
RARA	Retinoic Acid Receptor Alpha
RBC	Red blood cell count
RNA	Ribonucleic acid
RUNX1	Runt-related transcription factor 1
RUNX1T1	RUNX1 Translocation Partner 1
RUNX2	Runt-related transcription factor 2
RWDD4	RWD Domain Containing 4
S100A8	S100 Calcium Binding Protein A8
S100A9	S100 Calcium Binding Protein A9
sAML	secondary AML
SAT1	Spermidine/Spermine N1-Acetyltransferase 1
SELL	Selectin L
SF3B1	Splicing factor 3b
SIRPD	Signal Regulatory Protein Delta
SKCM	Skin cancer
SMYD3	SET And MYND Domain Containing 3
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SRSF2	Serine/arginine-rich splicing factor 2
STAT	Signal transducer and activator of transcription
STAT5B	Signal Transducer And Activator Of Transcription 5B
TAA	Tumor associated antigens
TBRG1	Transforming Growth Factor Beta Regulator 1
TET2	Ten-eleven translocation oncogene family member 2
TGF-ß	Transforming growth factor beta 1
TMCC2	Transmembrane And Coiled-Coil Domain Family 2
TP53	Tumor protein p53
TPO	Thrombopoietin
TTI1	TELO2 Interacting Protein 1
TYK2	Tyrosine kinase 2
U2AF1	U2 small nuclear RNA auxiliary factor 1

UBL7	Ubiquitin Like 7
UM	Uveal melanoma
UPD	Uniparental disomy
UTR	Untranslated region
UV	Uveal melanoma
VAF	Variant allele frequency
VWA7	Von Willebrand Factor A Domain Containing 7
WHO	World Health Organization
XCI	X chromosome inactivation
XIST	X-inactive specific transcript
XPO5	Exportin 5
ZFP36	ZFP36 Ring Finger Protein
ZFYVE27	Zinc Finger FYVE-Type Containing 27
ZNF397	Zinc Finger Protein 397
ZRSR2	Zinc finger, RNA-binding motif and serine/arginine rich 2

APPENDICES

CURRICULUM VITAE

Fiorella Schischlik

⦿ Untere Augartenstraße 9/14, 1020 Vienna, Austria
✉ fiorella.schischlik@gmail.com
☎ +43 680 4444391

EDUCATION

Medical University Vienna   Vienna, Austria
PhD candidate at CeMM under the supervision of Robert Kralovics   September 2012 - present

University of Applied Sciences 'Campus Wien'   Vienna, Austria
M.Sc. Eng., Bioinformatics   September 2010 - June 2012
- Master thesis (Boehringer Ingelheim): *Detection and annotation of genomic alterations in cancer cell lines*

University of Applied Sciences 'Technikum Wien'   Vienna, Austria
M.Sc. Eng and B.Sc. Eng., Biomedical Engineering   September 2005 - June 2011
- Master thesis (MFPL): *Origin and diversification of the glycoprotein hormone system*
- Bachelor thesis (Boehringer Ingelheim): *Nanobodies as novel agents for disease therapy*

Exchange semesters abroad
Linköping University (**Linköping, Sweden**)   September 2009 - January 2010
Universidad de Vigo (**Vigo, Spain**)   February 2007 - June 2007

CONFERENCES AND WORKSHOPS

ASH Conference. (Accepted Poster Presentation). *December 2018. San Diego, USA*

ISMB Conference. Poster Presentation: **Mutational Landscape of the Transcriptome in Patients with Hematological Malignancies.** *July 2017. Prague, Czech Republic.*

SAB (Scientific Advisory Board) Meeting. Oral Presentation. *February 2017. CeMM, Vienna, Austria*

ESH (European School of Hematology). Poster presentation: **Mutational Landscape of the Transcriptome in Patients with Chronic State and Transformed Myeloproliferative Neoplasms.** *October 2016. Estoril, Portugal.*

ASH Conference. Poster Presentation: **Fusion Gene Detection Using Whole Transcriptome Analysis in Patients with Chronic Myeloproliferative Neoplasms and Secondary Acute Myeloid Leukemia.** *December 2015. Orlando, USA.*

Cancer Genomics Workshop. *30 June - 04 July 2014. EBI, Cambridge, England.*

Cancer Genomics Conference. *November 2013. EMBL, Heidelberg, Germany.*

FELLOWSHIPS AND ADDITIONAL INFORMATION

PhD representative 2014: Organized together with PhD representatives and PhD students from EULife partner institute CRG in Barcelona, a joined 'Young Scientist Retreat' at Prein an der Rax.

Achievement Scholarship for study program Biomedical Engineering for the study year 2008/2009

Lower Austria Fellowship (Winter semester WS 2009/2010)

Erasmus Fellowship for Vigo University, Spain (Summer semester SS 2007)

PUBLICATIONS *first author

Schischlik F*, Jäger R, Rosbrock F, Hug E, Schuster M, Holly R, Milosevic Feenstra JD, Bogner E, Gisslinger B, Schalling M, Rumi E, Pietra D, Fischer G, Faé I, Vulliard L, Menche J, Meggendorfer M, Stengel A, Haferlach T, Bock C, Cazzola M, Gisslinger H, Kralovics R. **Mutational landscape of the transcriptome offers Applications for Immunotherapy of Myeloproliferative Neoplasms.** *(manuscript submitted 10-2018 to Blood; currently under review)*

Bigenzahn JW, Collu GM, Kartnig F, Pieraks M, Vladimer GI, Heinz LX, Sedlyarov V, Schischlik F, Fauster A, Rebsamen M, Parapatics K, Blomen VA, Müller AC, Winter GE, Kralovics R, Brummelkamp TR, Mlodzik M, Superti-Furga G. **LZTR1 is a regulator of RAS ubiquitination and signaling.** *Science. 11-2018*

Fauster A, Rebsamen M, Willmann KL, César-Razquin A, Girardi E, Bigenzahn JW, <u>Schischlik F</u>, Scorzoni S, Bruckner M, Konecka J, Hörmann K, Heinz LX, Boztug K, Superti-Furga G. **Systematic genetic mapping of necroptosis identifies SLC39A7 as modulator of death receptor trafficking.** *Cell Death and Differentiation.* 2018.

<u>Schischlik F*</u>, Kralovics R. **Mutations in myeloproliferative neoplasms–their significance and clinical use.** *Expert Review of Hematology.* 2017

Moder M, Velimezi G, Owusu M, Mazouzi A, Wiedner M, Ferreira da Silva J, Robinson-Garcia L, <u>Schischlik F</u>, Slavkovsky R, Kralovics R, Schuster M, Bock C, Ideker T, Jackson SP, Menche J, Loizou JI. **Parallel genome-wide screens identify synthetic viable interactions between the BLM helicase complex and Fanconi anemia.** *Nature Communications.* 2017

Kapralova K, Horvathova M, Pecquet C, Fialova Kucerova J, Pospisilova D, Leroy E, Kralova B, Milosevic Feenstra JD, <u>Schischlik F</u>, Kralovics R, Constantinescu SN, Divoky V. **Cooperation of germ line JAK2 mutations E846D and R1063H in hereditary erythrocytosis with megakaryocytic atypia.** *Blood.* 2016

Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD, Them NC, Berg T, Gisslinger B, Pietra D, Chen D, Vladimer GI, Bagienski K, Milanesi C, Casetti IC, Sant'Antonio E, Ferretti V, Elena C, <u>Schischlik F</u>, Cleary C, Six M, Schalling M, Schönegger A, Bock C, Malcovati L, Pascutto C, Superti-Furga G, Cazzola M, Kralovics R. **Somatic mutations of calreticulin in myeloproliferative neoplasms.** *New England Journal of Medicine.* 2013.

Bürckstümmer T, Banning C, Hainzl P, Schobesberger R, Kerzendorfer C, Pauler FM, Chen D, Them N, <u>Schischlik F</u>, Rebsamen M, Smida M, Fece de la Cruz F, Lapao A, Liszt M, Eizinger B, Guenzl PM, Blomen VA, Konopka T, Gapp B, Parapatics K, Maier B, Stöckl J, Fischl W, Salic S, Taba Casari MR, Knapp S, Bennett KL, Bock C, Colinge J, Kralovics R, Ammerer G, Casari G, Brummelkamp TR, Superti-Furga G, Nijman SM. **A reversible gene trap collection empowers haploid genetics in human cells.** *Nature Methods.* 2013.

SUPPLEMENTARY TABLE 1 – PATIENT CLINICAL DATA

unique.patient.id	upn.clon	diagnosis.final	genotype.detail	genotype.reduced	sf3b1.hotspot.aa.change	jak2.burden	calr.burden	sex	patient.age.in.years	survival.in.months	Hb.at.diagn	WBC.at.diagn	Plt.at.diagn	previous.thrombosis	thrombotic.events.after.diagn
C001A	CA001	Control	wt	wt	wt	0	0	F	28,4	NA	NA	NA	NA	NA	NA
C002A	CA002	Control	wt	wt	wt	0	0	M	NA	NA	NA	NA	NA	NA	NA
C003A	CA003	Control	wt	wt	wt	0	0	M	NA	NA	NA	NA	NA	NA	NA
C004A	CA004	Control	wt	wt	wt	0	0	F	41	NA	NA	NA	NA	NA	NA
C005A	CA005	Control	wt	wt	wt	0	0	F	60	NA	NA	NA	NA	NA	NA
C006A	CA006	Control	wt	wt	wt	0	0	M	44	NA	NA	NA	NA	NA	NA
C007A	CA007	Control	wt	wt	wt	0	0	M	NA	NA	NA	NA	NA	NA	NA
C008A	CA008	Control	wt	wt	wt	0	0	F	42	NA	NA	NA	NA	NA	NA
C009A	CA010	Control	wt	wt	wt	0	0	M	40	NA	NA	NA	NA	NA	NA
C010A	CA012	Control	wt	wt	wt	0	0	F	27,4	NA	NA	NA	NA	NA	NA
C011A	CA013	Control	wt	wt	wt	0	0	M	49,5	NA	NA	NA	NA	NA	NA
C012A	CA015	Control	wt	wt	wt	0	0	F	38,5	NA	NA	NA	NA	NA	NA
C013A	CA016	Control	wt	wt	wt	0	0	F	25,9	NA	NA	NA	NA	NA	NA
C014A	CA017	Control	wt	wt	wt	0	0	M	23,9	NA	NA	NA	NA	NA	NA
C015A	CA018	Control	wt	wt	wt	0	0	M	26,9	NA	NA	NA	NA	NA	NA
P004A	PA004	ET	JAK2-V617F	JAK2-V617F	wt	30	0	F	59,9	129	11,6	14,3	905	yes	no
P005C	PC005	ET	CALR-ins5	CALR-type2	wt	0	43,8	F	64,9	220	13,2	9,99	643	no	no
P008C	PC008	ET	Triple-neg	Triple-neg	wt	0	0	F	78	165	14,6	5,8	624	no	no
P010A	PA010	ET	JAK2-V617F	JAK2-V617F	wt	12	0	F	50,4	273	14,7	7,1	677	no	no
P011D	PD011	ET	JAK2-V617F	JAK2-V617F	wt	41	0	F	59,8	386	13,9	8,9	1550	yes	no
P014A	PA014	ET	JAK2-V617F	JAK2-V617F	wt	42	0	F	66,1	213	15,2	9,7	725	yes	no
P016A	PA016	ET	JAK2-V617F	JAK2-V617F	wt	82	0	F	84,6	175	14,2	11,3	1050	no	yes
P018A	PA018	ET	JAK2-V617F	JAK2-V617F	wt	26	0	M	51,5	206	13,4	10	770	yes	yes
P022A	PA022	ET	CALR-ins5	CALR-type2	wt	0	49,6	F	67,9	156	11,7	6,82	839	no	no
P023A	PA023	ET	CALR-del52	CALR-type1	wt	0	40,6	F	67,5	207	13,6	7,12	906	no	no
P028B	PB028	ET	JAK2-V617F	JAK2-V617F	wt	33	0	M	52,8	159	15,6	11,3	1122	yes	no
P030C	PC030	ET	CALR-ins5	CALR-type2	wt	0	46,9	M	64,4	275	14,6	6,87	1160	no	yes
P031A	PA031	ET	JAK2-V617F	JAK2-V617F	wt	42	0	M	77,3	65	12,6	7,8	491	yes	no
P035C	PC035	ET	Triple-neg	Triple-neg	wt	0	0	F	56	278	14	15,65	1320	no	no
P036A	PA036	ET	JAK2-V617F	JAK2-V617F	wt	52	0	F	80,4	217	13,9	10,5	1117	yes	no
P045A	PA045	ET	CALR-ins5	CALR-type2	wt	0	49,7	F	69,3	323	15,2	12,1	1600	no	no
P047A	PA032	ET	CALR-ins5	CALR-type2	wt	0	43,9	F	72,7	192	14,6	5,35	769	no	no
P047B	PB047	ET	CALR-ins5	CALR-type2	wt	0	43,9	F	72,7	192	14,6	5,35	769	no	no
P049A	PA049	ET	CALR-Type10	CALR-type2	wt	0	47,9	F	53	262	13,6	14,45	829	no	no
P050B	PB050	ET	JAK2-V617F	JAK2-V617F	wt	43	0	F	73,2	219	14,8	6,69	653	no	no
P052A	PA052	ET	JAK2-V617F	JAK2-V617F	wt	35	0	M	70,7	185	13,4	8,34	531	yes	no
P057A	PA057	ET	CALR-ins5	CALR-type2	wt	0	50,2	F	67,2	161	14,4	9,4	889	no	no
P061A	PA061	ET	CALR-del52	CALR-type1	wt	0	34,6	M	71,7	135	12,3	9,8	786	yes	no
P062A	PA062	ET	Triple-neg	Triple-neg	wt	0	0	F	58,1	141	11,4	6,65	580	no	no
P066A	PA066	ET	CALR-del52	CALR-type1	wt	0	54	M	63,6	133	14,2	7,3	838	yes	no
P068A	PA068	ET	CALR-del52	CALR-type1	wt	0	49,5	M	43,8	133	14,6	13,5	1306	yes	no
P076B	PB075	ET	JAK2-V617F	JAK2-V617F	wt	18	0	F	72,5	121	13,7	7,9	774	no	no
P079C	PC078	ET	JAK2-V617F	JAK2-V617F	wt	70	0	M	75,2	166	15,8	11,3	436	yes	no
P081B	PB080	ET	JAK2-V617F	JAK2-V617F	wt	4	0	F	83,8	302	14,5	6,64	624	no	no
P083B	PB082	ET	JAK2-V617F	JAK2-V617F	wt	48	0	F	79,7	67	13,3	8,98	642	yes	yes
P097A	PA095	ET	Triple-neg	Triple-neg	wt	0	0	M	53	78	15,6	13,29	668	no	no
P101A	PA099	ET	JAK2-V617F	JAK2-V617F	wt	36	0	F	42,9	98	17,6	9,96	951	yes	no
P103A	PA101	ET	JAK2-V617F	JAK2-V617F	wt	40	0	F	44,2	97	12,6	6,4	540	no	no
P001B	PB001	PMF	CALR-ins5	CALR-type2	wt	0	47,7	M	86	384	12,1	4,01	2530	yes	no
P003B	PB003	PMF	CALR-del52	CALR-type1	wt	0	54,3	F	65	149	12,9	7,25	1000	no	no
P006C	PC006	PMF	CALR-del52	CALR-type1	wt	0	67,6	M	62,4	195	14,1	9,5	1249	no	no
P009A	PA009	PMF	JAK2-V617F	JAK2-V617F	K666N	48	0	F	85,1	111	13,1	14,98	1063	no	yes
P013B	PB013	PMF	CALR-del52	CALR-type1	wt	0	54,3	F	83,2	216	11	6,01	114	no	no
P015A	PA015	PMF	CALR-del52	CALR-type1	wt	0	54,8	F	78,6	43	11,7	8,35	964	no	no
P019B	PB019	PMF	CALR-del52	CALR-type1	wt	0	59,1	M	85,8	189	10,6	7,7	568	no	yes
P020A	PA020	PMF	CALR-ins5	CALR-type2	wt	0	45,8	F	62,3	187	12,2	8,9	1182	no	no
P021A	PA021	PMF	CALR-ins5	CALR-type2	wt	0	50,5	F	60,2	230	13,3	13,1	1161	no	yes
P026A	PA026	PMF	JAK2-V617F	JAK2-V617F	wt	51	0	F	72,7	141	13,6	9,38	809	yes	no
P027A	PA027	PMF	CALR-del52	CALR-type1	wt	0	49,8	F	81,9	147	11,2	11,41	1226	no	yes
P029A	PA029	PMF	CALR-del52	CALR-type1	wt	0	51,2	M	66,7	248	10,2	5,3	274	no	no
P032A	PA032	PMF	CALR-ins5	CALR-type2	wt	0	51,8	M	80,7	281	12,5	5,82	574	no	no
P033C	PC033	PMF	CALR-Type33	CALR-type2	wt	0	43,2	F	71,5	236	12,2	11,5	1850	no	yes
P034A	PA034	PMF	JAK2-V617F	JAK2-V617F	K700E	59	0	F	54,7	127	10,4	7,68	247	no	no
P037B	PB037	PMF	CALR-Type12	CALR-type2	wt	0	43,8	M	74,8	181	14,7	9,29	1145	no	no

unique.patient.id	upn.clon	diagnosis.final	genotype.detail	genotype.reduced	sf3b1.hotspot.aa.change	jak2.burden	calr.burden	sex	patient.age.in.years	survival.in.months	Hb.at.diagn	WBC.at.diagn	Plt.at.diagn	previous.thrombosis	thrombotic.events.after.diagn
P039C	PA039	PMF	CALR-del52	CALR-type1	wt	0	65,1	F	56,5	202	NA	NA	NA	NA	NA
P040B	PB040	PMF	CALR-del52	CALR-type1	K666N	0	70,2	M	72,8	191	14,2	9,51	600	no	no
P041C	PA041	PMF	CALR-ins5	CALR-type2	wt	0	NA	F	74,6	132	12,3	10,35	546	no	no
P042A	PA042	PMF	JAK2-V617F	JAK2-V617F	wt	44	0	M	58,8	192	13,2	15,57	679	no	no
P044A	PA044	PMF	JAK2-V617F	JAK2-V617F	wt	62	0	F	78,3	142	12,3	19,01	872	yes	no
P046C	PC046	PMF	JAK2-V617F	JAK2-V617F	wt	10	0	M	72,7	93	14,1	15,49	635	yes	yes
P048A	PA048	PMF	CALR-ins5	CALR-type2	K666T	0	NA	M	53,4	241	14,5	9,98	716	no	no
P051D	PD051	PMF	CALR-Type14	CALR-type1	wt	0	NA	M	79	261	13	7,8	682	no	no
P053A	PA053	PMF	JAK2-V617F	JAK2-V617F	wt	46	0	F	54,2	137	14	8,22	646	no	no
P055C	PC055	PMF	CALR-ins5	CALR-type2	K700E	0	49,5	F	61,8	293	10,4	5,7	604	no	no
P056A	PA056	PMF	JAK2-V617F	JAK2-V617F	wt	40	0	F	43,7	166	13,7	9,2	803	no	no
P058A	PA058	PMF	JAK2-V617F	JAK2-V617F	K700E	59	0	F	77,6	56	10,7	9,8	370	no	yes
P059B	PB059	PMF	JAK2-V617F	JAK2-V617F	wt	32	0	M	47,2	170	16	10,2	810	no	no
P060A	PA060	PMF	CALR-del52	CALR-type1	wt	0	50,1	M	50,7	NA	14	8,19	874	no	no
P060B	PB060	PMF	CALR-del52	CALR-type1	wt	0	50	M	56,3	144	14	8,19	874	no	no
P064A	PA064	PMF	JAK2-V617F	JAK2-V617F	K666R	21	0	M	80,5	159	13,2	6,9	621	no	yes
P065C	PC065	PMF	CALR-ins5	CALR-type2	wt	0	46,2	F	49,7	380	11,8	20,6	1440	no	no
P067A	PA067	PMF	JAK2-V617F	JAK2-V617F	wt	44	0	M	71,3	64	9	3,11	393	yes	no
P069A	PA069	PMF	JAK2-V617F	JAK2-V617F	K700E	62	0	M	62,4	87	8,8	5,33	161	no	no
P073A	PA072	PMF	JAK2-V617F	JAK2-V617F	wt	29	0	F	50,1	196	14,7	7,7	1095	no	no
P074B	PB073	PMF	JAK2-V617F	JAK2-V617F	wt	22	0	M	70	108	15,1	14,39	1322	no	no
P075A	PA074	PMF	JAK2-V617F	JAK2-V617F	wt	53	0	M	68,1	147	12,7	39,26	199	no	no
P077C	PC076	PMF	CALR-del52	CALR-type1	wt	0	52,9	M	56,4	397	15,7	7,9	785	no	no
P078A	PA077	PMF	JAK2-V617F	JAK2-V617F	wt	52	0	M	69,2	259	14,6	8,1	524	no	yes
P080B	PB079	PMF	CALR-del52	CALR-type1	wt	0	50,7	F	37	128	11,5	8,14	977	no	no
P082B	PB081	PMF	CALR-del52	CALR-type1	wt	0	47,1	F	39,8	110	11,5	5,82	243	no	no
P085A	PA084	PMF	CALR-Type19	CALR-type1	wt	0	51,4	F	44	98	13,4	5,78	824	no	yes
P087A	PA086	PMF	JAK2-V617F	JAK2-V617F	wt	41	0	M	55,4	27	6,6	4,8	125	no	no
P089B	PB088	PMF	JAK2-V617F	JAK2-V617F	wt	65	0	F	45	89	13,5	7,95	326	yes	no
P090A	PA089	PMF	JAK2-V617F	JAK2-V617F	wt	34	0	F	49,5	99	12,9	10,6	850	no	no
P093A	PA091	PMF	JAK2-V617F	JAK2-V617F	wt	46	0	F	73,2	75	11,9	8,5	524	no	no
P094B	PB092	PMF	JAK2-V617F	JAK2-V617F	wt	82	0	M	72	75	13,6	9,17	612	yes	yes
P096A	PA094	PMF	CALR-ins5	CALR-type2	wt	0	44,4	F	61,3	181	12,3	10,3	567	no	no
P098A	PA086	PMF	CALR-del52	CALR-type1	wt	0	21,29	F	69,1	56	13,1	13,39	1131	no	no
P099A	PA097	PMF	CALR-del52	CALR-type1	wt	0	53,9	M	70,2	75	15	9,81	536	no	no
P100A	PA098	PMF	JAK2-V617F	JAK2-V617F	wt	51	0	M	73,6	103	11,8	6,82	220	no	no
P104A	PA102	PMF	JAK2-V617F	JAK2-V617F	wt	20	0	F	64	72	13,8	9,48	689	no	no
P105A	PA103	PMF	JAK2-V617F	JAK2-V617F	wt	34	0	F	70,5	70	15,2	12,05	566	no	no
P106A	PA104	PMF	CALR-Type7	CALR-type2	wt	0	56,9	M	84,6	73	10,4	8,09	402	yes	no
P107A	PA105	PMF	JAK2-V617F	JAK2-V617F	wt	81	0	M	72,8	57	11,2	12,3	1262	no	no
P002A	PA002	PV	JAK2-V617F	JAK2-V617F	wt	88	0	F	84,6	67	NA	NA	NA	NA	NA
P007B	PA007	PV	JAK2-V617F	JAK2-V617F	wt	87	0	M	64,5	200	NA	NA	NA	NA	NA
P012B	PB012	PV	JAK2-ex12	JAK2-ex12	wt	41	0	F	67	95	NA	NA	NA	NA	NA
P017B	PB017	PV	JAK2-V617F	JAK2-V617F	wt	100	0	F	50,9	262	NA	NA	NA	NA	NA
P024A	PA024	PV	JAK2-V617F	JAK2-V617F	wt	100	0	M	64,7	245	NA	NA	NA	NA	NA
P025A	PA025	PV	JAK2-V617F	JAK2-V617F	wt	98	0	F	48,1	40	NA	NA	NA	NA	NA
P038A	PA038	PV	JAK2-V617F	JAK2-V617F	wt	63	0	M	67,7	212	NA	NA	NA	NA	NA
P043C	PC043	PV	JAK2-V617F	JAK2-V617F	wt	49	0	F	60,9	207	NA	NA	NA	NA	NA
P063A	PA063	PV	JAK2-V617F	JAK2-V617F	wt	99	0	M	68	326	NA	NA	NA	NA	NA
P071A	PA070	PV	JAK2-V617F	JAK2-V617F	wt	82	0	F	37,7	173	NA	NA	NA	NA	NA
P072A	PA071	PV	JAK2-V617F	JAK2-V617F	wt	66	0	F	67,7	40	NA	NA	NA	NA	NA
P084A	PA083	PV	JAK2-V617F	JAK2-V617F	wt	91	0	M	72,2	93	NA	NA	NA	NA	NA
P086A	PA085	PV	JAK2-V617F	JAK2-V617F	wt	94	0	M	69,7	103	NA	NA	NA	NA	NA
P088A	PA087	PV	JAK2-V617F	JAK2-V617F	wt	95	0	M	63,7	172	NA	NA	NA	NA	NA
P091A	PA090	PV	JAK2-V617F	JAK2-V617F	wt	95	0	F	76,1	76	NA	NA	NA	NA	NA
P095A	PA093	PV	JAK2-V617F	JAK2-V617F	wt	64	0	F	72	82	NA	NA	NA	NA	NA
P102A	PA100	PV	JAK2-V617F	JAK2-V617F	wt	100	0	M	60,1	71	NA	NA	NA	NA	NA
P108A	PA106	sAML	JAK2-V617F	JAK2-V617F	wt	49	0	F	85,2	NA	NA	NA	NA	NA	NA
P109A	PA107	sAML	JAK2-V617F	JAK2-V617F	wt	56	0	M	73,4	NA	NA	NA	NA	NA	NA
P110A	PA108	sAML	JAK2-V617F	JAK2-V617F	wt	56	0	F	78,8	NA	NA	NA	NA	NA	NA
P111A	PA109	sAML	JAK2-V617F	JAK2-V617F	wt	1,6	0	M	71,3	NA	NA	NA	NA	NA	NA
P112A	PA110	sAML	JAK2-V617F	JAK2-V617F	wt	94	0	M	56,6	NA	NA	NA	NA	NA	NA
P113A	PA111	sAML	JAK2-V617F	JAK2-V617F	wt	71	0	M	74,3	NA	NA	NA	NA	NA	NA
P114A	PA112	sAML	JAK2-V617F	JAK2-V617F	wt	65	0	F	73,2	NA	NA	NA	NA	NA	NA
P115A	PA113	sAML	JAK2-V617F	JAK2-V617F	wt	98	0	F	73,9	NA	NA	NA	NA	NA	NA
P116A	PA114	sAML	MPL	MPL	wt	0	0	M	82,8	NA	NA	NA	NA	NA	NA